심리학석사학위논문

Regularizing Structural Equation Models via the Lasso : Generalizability and Reproducibility Issues

2016 년 8 월

서울대학교 대학원

심리학과 계량심리 전공

강 인 한

심리학석사학위논문

Regularizing Structural Equation Models via the Lasso : Generalizability and Reproducibility Issues

2016 년 8 월

서울대학교 대학원

심리학과 계량심리 전공

강 인 한

Regularizing Structural Equation Models via the Lasso
: Generalizability and Reproducibility Issues

지도교수 김 청 택

이 논문을 심리학석사 학위논문으로 제출함

2016 년 6 월

서울대학교 대학원

심리학과 계량심리 전공

강 인 한

강인한의 심리학석사 학위논문을 인준함

2016 년 6 월

위 원 장 _____

부위원장 _____

위 원 _____

Abstract

Generalizability and Reproducibility of research have become one of the main topics in current psychology. Previous discussions on the issue have focused on the 'Experimental/Procedural' aspect such as incentive structure for researchers, violation in conducting an experiment, selective reporting, etc.

However, sometimes statistical methods which are widely used in psychology have properties that undermine the generalizabilty of research results. The present thesis approaches the reproducibility problem based on this 'Analytical/Statistical' aspect. For this purpose, we studied a method for improving the 'Structural Equation Modeling(SEM)', one of dominant statistical models in psychology. The main focus of this study is implementing L_1-regularization, or 'Lasso', to SEM. With this method, the result will enjoy less variability of estimation than the existing Maximum Likelihood method.

First of all, the present thesis discusses some indices including 'Overall Discrepancy(OD)' and 'Mean Squared Error(MSE)' as criteria which indicate the generalizability and reproducibility of analysis results. Bayesian Lasso SEM, one of the previous attempts, is also covered with some fundamental issues. Furthermore, an algorithm for regularizing SEM via the Lasso is derived and examined by several simulation studies.

The study is carried out using Factor Analysis Model and Structural Equation Modeling, while adding several misspecified parameters. The purpose of this approach is to test Lasso SEM's complete shrinkage ability, which is able to detect and remove unnecessary parameters from the original model so that the method yields the result close to the true population-generating process.

It is also investigated whether Lasso can improve generalizability and reproducibility by observing and comparing OD and MSE. The simulation deals with various conditions including model error, sample sizes, and magnitudes of covariance matrix, in order to examine in which condition Lasso SEM yields better results than the Maximum Likelihood Estimation.

The result reveals that Lasso SEM works well in various conditions; it improves generalizability indices, detects and removes misspecified parameters in the original model. However, the performance depends on the conditions, which implies that the Lasso SEM should be applied with careful scrutiny on characteristics of practical data. Especially, the model error, one of the component affecting the data-generating process, has turned out to be the most influential factor that hinders proper function of the Lasso SEM. We suggest modifying the optimization of Lasso SEM, which is currently rely upon the value of OD, or its cross-validation estimate. The improvement can be achieved by replacing criteria or objective function in the optimization procedure. This will minimize problems including those generated from the model error.

A correlation analysis shows that 'Sample Discrepancy', which is a criterion of the existing estimation method, and goodness of model fit indices widely used in SEM have considerably low correlations with OD. This outcome implies the SEM result obtained by the original method may be hard to be generalized to other independent samples including the future data, and the phenomenon that researchers are interested in.

Keywords: Reproducibility, Generalizability, Structural Equation Modeling, Factor Analysis, Regularization, Lasso, Overall Discrepancy, Mean Squared Error

Student Number: 2014-20229

Contents

List of Figures

List of Tables

Introduction

Reproducibility Issues in Psychological Researches

'Reproducibility' is one of terms that represent a new era of recent scientific psychology. Only if effects proposed and verified in psychology are replicated in subsequent studies under the same conditions and experimental procedures, they finally gain scientific values(Schmidt, 2009; Koole & Lakens, 2012; Open Science collaboration, 2012, 2014, 2015). Without reproducibility, the results of scientific researches might be depreciated as only the meaningless by-products from random noise.

As a matter of fact, many a scholar had discussed this issue for decades in slightly different contexts of power analysis, effect size, meta analysis and so on(Cohen, 1962; Sedlmeier & Gigerenzer, 1989; Rossi, 1990; Cohen, 1992; Cohen, 1994). However, unfortunately, even though these discussions may have cautioned psychologists in some degree, they had not received as much attention as they should have.

Recently, this issue arise sensationally with the terminology 'Reproducibility Crisis', raising a question whether psychological researches are able to be replicated. And now, the reproducibility issue is regarded as a paradigm-shift, at the same time a problem which may undermine the name of psychology as a field of science.

Regarding the issue, a number of researchers started to discuss and investigate the reproducibility of researches in psychology(Neuliep & Crandall, 1990; Schmidt & Hunter, 1996; Klein, Brown & Lysyk, 2000; Ioannidis, 2005; Schmidt, 2009; LeBel & Paunonen, 2011; Simmons, Nelson & Simonsohn, 2011; Francis, 2012; Hartshorne & Schachner, 2012; Koole & Lakens, 2012; Makel, Plucker & Hegarty, 2012; Nosek & Bar-Anan, 2012; Nosek, Spies & Motyl, 2012; Open Science Collaboration, 2012, 2014, 2015; Errington et al., 2014; Klein et al., 2014;). Furthermore, Some of them launched large-scale collaborations to deal with the issue. Especially, 'Many Labs Replication Project'(Klein et al., 2014) and 'Open Science Collaboration Reproducibility Project'(Open Science Collaboration, 2015), henceforth abbreviated as ML2014 and OSC2015 respectively, are attracting huge attention as seminal researches on the reproducibility issue. These two collaborations conducted intensive studies regarding the replication of previous researches published in psychological journals.

ML2014 attempted to replicate 13 well-known effects in psychology, with 36 samples that have different characteristics such as countries where the sampling is occurred(US or not), and methods by which sampling is conducted(laboratory or online). Previous experiments were re-conducted by each of 36 samples in their database, so that each effect was tested 36 times. Based on the effect size Cohen's d and its 99% confidence interval, they reported that 10 out of 13 effects were well-reproduced. Another one effect was weakly-replicated, and the other two effects could not be said to be recreated.

OSC2015 implemented a different approach on the same issue. Above all, the collaboration established a protocol regarding how to select journals and researches to be replicated, how to allocate their academic volunteers to each of chosen researches, and how to conduct the replication researches. Then, they attempted to regenerate the results of 100 researches published in *Psychological*

Science, Journal of Personality and Social Psychology, and *Journal of Experimental Psychology: Learning, Memory, and Cognition*. Multiple criteria to judge whether the research results are reproduced were statistical significance and p-values, effect sizes, subjective assessment by their own replication teams, and meta-analysis of effect sizes. Comparing the results from original and replication researches, the project arrived at a different conclusion from that of ML2014; Among the 100 researches, only 36-47% of original researches are reproduced depending on various criteria. Also, The mean effect size declined to a not inconsiderable extent, and only 39% studies were judged as being replicated successfully by their replication teams.

Rather speaking about the validity of the two projects and their conclusions, we can find valuable implications from them. Especially, thanks to their contribution, the 'reproducibility' issue has become the chief topic of recent studies in psychology. What should be noted is that the research results, and even the methods of these projects failed to reach a general consensus. This implies the reproducibility issue in psychology will face intense controversy and debate henceforth. Therefore, it seems psychologists' endeavor to conquer the reproducibility issue has just started.

Analytical/Statistical Approach to Reproducibility Issues

Concerning researches of ML2014, OSC2015, and other researches dealing with the issue of reproducibility, we are able to reach a common ground in approaching the problem. The previous researchers focused on the 'Experimental/Procedural' aspect in analyzing and discussing the reproducibility. This term is intended to refer that so far discussions on replication and reproducibility have focused on the habitual practices and customs of researchers,

from starting their studies to publishing the articles. These practices are mainly caused by the incentive structure emphasizing the novelty rather than accuracy(Klein, Brown & Lysyk, 2000; Open Science Collaboration, 2012, 2014, 2015; Hartshorne & Schachner, 2012; Nosek, Spies & Motyl, 2012; Errington et al., 2014). For example, among these customs are publication or editorial bias(Neuliep & Crandall, 1990; Schmidt, 2009; Francis, 2012; Hartshorne & Schachner, 2012; Koole & Lakens, 2012; Errington et al., 2014), selective reporting(Francis, 2012; Open Science Collaboration, 2014, 2015), selective analysis(Koole & Lakens, 2012; Open Science Collaboration, 2015), undisclosed and inadequate flexibility in designs, data collection, and analysis(Ioannidis, 2005; Simmons, Nelson & Simonsohn, 2011; Hartshorne & Schachner, 2012), violation in conducting an experiment and statistical analysis(Francis, 2012), experimenter bias and file-drawer problem(Makel, Plucker & Hegarty, 2012), data fabrication(Koole & Lakens, 2012), low power and effect size(Sedlmeier & Gigerenzer, 1989; Ioannidis, 2005; Hartshorne & Schachner, 2012), implementation of measurements turned out to contain a huge amount of random measurement error(Schmidt & Hunter, 1996; LeBel & Paunonen, 2011), incomplete reporting of experimental conditions required to obtain the results(Errington et al., 2014; Open Source Collaboration, 2015), reluctance to share data and rapid loss of data(Errington et al., 2014), and narrative(philosophical, theoretical, and rhetorical) approach to scientific publication in psychology(Koole & Laken, 2012). Those practices are not independent from each other; they are conceptually overlapped, and they interact with one another as some of them may cause others, and again they bring about other practices. Furthermore, in doing so, they aggravate the reproducibility problem.

Also, discussion on remedies for these habitual practices and customs in scientific communication have also focused on the 'Experimental/Procedural' as-

pect. Notable examples are six changes suggested by Nosek and Bar-Anan(2012), three incentives presented by Koole and Lakens(2012), six requirements for researchers and four guidelines for reviewers provided in Simmons, Nelson, and Simonsohn(2011), '*Replication Links*' and '*Replication Tracker*' proposed by Hartshorne and Schachner(2012).

However, a different point of view, which we call the 'Analytical/Statistical' aspect, need to be considered in order to expand the discussion on reproducibility issue. The terms, 'analytical' and 'statistical' are easily noticed in previous articles dealing with the same topic. However, we intend to indicate a different approach by these terms. At times, some statistical methods exploited to prove researchers' theory can be the issue. Even though their statistical property is not flawless, some methods may contain several limitations on application.

In order to understand this point, we shall discuss about the estimation problem. In statistics, we construct an 'estimator' $\hat{\theta}$ to estimate its parameter θ and study its statistical property to examine whether $\hat{\theta}$ is appropriate to estimate θ. One of the dominant indices representing a goodness of estimator is the 'Mean squared Error', or MSE. It is well known that this index is equal to the sum of 'bias' squared and 'variance' of estimator.

In the early period of statistics, most estimators were developed focusing on reduction of bias. That is, we prioritized 'Unbiased' estimators, whose expected value is equal to the target parameter. Among these kind of estimators, the one that has the minimum variance is chosen by statisticians. 'Ordinary Least Squares' and 'Maximum Likelihood Estimation' have been dominant in statistics since they are able to produce unbiased estimator in many(but not always) cases.

However, when the minimum variance of the unbiased estimator is not

small enough, this estimator may yield unstable results. This menas, when its estimate is computed with different samples, the results may vary remarkably. With this estimator, a result from the given sample cannot be 'generalized' to other independent samples including the one obtained in the future for predicting the phenomenon. In other words, this estimator cannot generate a reproducible result.

Regarding the discussion above, some statisticians started to recommend the 'Biased' estimator with less variance than minimum variance of the unbiased estimator. Notable examples of the biased estimator are James-Stein Estimator, and Regularization such as Ridge, Lasso, and Elastic net, which will be described in the present thesis. However, even though these methods are able to produce less variance, it doesn't mean that they are better than existing unbiased estimators if they suffer from large bias; Despite the fact that their results are stable, this results might not be accurate. Therefore, MSE, the overall index which encompasses both bias and variance, can be used as a criterion of generalizability and reproducibility of the analysis results. Those estimators such as James-Stein estimator and regularization methods have already been proven to produce less variance with not-too-large bias. Therefore, they can reduce MSE as well.

In this regard, stepping away from the old convention which focuses on 'unbiasedness', statisticians have been preparing for the methods which are able to yield more generalizable and reproducible results. In fact, estimation under this spirit is one of the mainstream of modern statistics.

Generalizability in Structural Equation Modeling

Despite the change of mainstream in statistics, the philosophy and its resulting methods have not been expanded to the field of psychology. In psychology, researchers take advantage of various statistical methods in studying on cognition, emotion, and behavior of human being. One of the dominant statistical models in psychology is the 'Structural Equation Modeling'. This model has long been favored since it enables psychologists to take unobservable 'psychological constructs' into account through the concept of 'latent variable', or 'factor'.

However, generalizability or reproducibility of results obtained from the model has not been thoroughly discussed. Even though SEM has numerous indices for evaluating goodness of model fit, the vast majority of these criteria do not take the issue into consideration. In point of fact, fit indices are not related directly to the notion of MSE, bias, and variance. Since SEM pursues explaining the underlying covariance structure among the variables, rather than investigating the estimation of each parameter, those fit indices are defined in connection with 'discrepancy' between several matrices. Among these discrepancies are Overall Discrepancy(OD), Discrepancy due to Approximation(DA), Discrepancy due to Estimation(DE), and Sample Discrepancy(SD). Details on these discrepancies and their conceptual relationships with MSE, bias, and variance will be described in Chapter 1. What should be noted here is that i) OD is conceptually related to MSE and can be regarded as an index for generalizability, and ii) those discrepancies, except SD which is an index defined based on the current sample and used in practical SEM estimation, are not observable since their computation needs matrices defined in population level. Therefore, even though SEM has a good criterion that enable us to figure

out generalizability/reproducibility of the model, it is impossible to observe and investigate it. Moreover, goodness of fit indices widely used in SEM are usually defined based on their relationships with SD, which is the only observable discrepancy but has no concern with replicable results. Hence, the generalizability and reproducibility issue on SEM results has not been at the centre of attention until recently.

However, considering the recent reproducibility crisis in psychology, all the statistical methods used in the field should be re-validated or improved with respect to their ability to produce a generalizable and reproducible result. Modern trends in statistics regarding biased but less variable estimation can be an effective tool to achieve this goal. For SEM(In fact, for all the traditional statistical methods), the 'Bayesian' approach can be an alternative for the existing estimation method. This approach conducts the estimation combining likelihood with the 'prior' densities of interested parameters. This distributions represent our knowledge, reasonable expectation, or obtained results in the previous researches which give information regarding those parameters. And this combination yields the 'posterior' distributions, which include comprehensive information on the corresponding parameters. Since this procedure adds prior information to likelihood, which is the objective function of Maximum Likelihood Estimation, it produces a different outcome from the ML results.

In fact, a Bayesian approach is highly connected with regularization in that it gives constraint to parameter space using the prior distribution. In this regard, Bayesian SEM seems to be able to yield more reproducible results than ML with appropriate priors. Furthermore, there has been endeavors to apply existing regularization methods such as Lasso by means of Bayesian framework(Park & Casella, 2008). One of those results, which is called 'BLasso', can

be employed in order to achieve our goal with more direct and intense regularization than ordinary Bayesian SEM. Even though these Bayesian approaches were expanded to SEM only recently(Lee, 2007; Guo et al., 2012) and have not received much attention from SEM users, they deserve special emphasis as a remedy for the reproducibility issue.

In the present thesis, however, we claims that Bayesian SEM and BLasso SEM cannot be the best choice for generalizability of the model. This argument is grounded on the fact that Bayesian statistical methods derive the posterior distributions even for the misspecified parameters which we want to remove by Lasso-type regularization. In other words, Bayesian approach is not able to delete these nuisance parameters and produce more sparse and parsimonious model, which is what Lasso does for improving generalizability and reproducibility.

In this perspective, we shall suggest another approach for improving the estimation of the structural equation modeling. For this purpose, Lasso, or 'Least Absolute Shrinkage and Selection Operator', is applied to the SEM in a direct and straightforward way, not in the Bayesian framework. As described, Lasso is one of the regularization methods in statistics which conducts simultaneous estimation and model selection. This method is known as being able to improve generalizability of the results obtained from linear regression models, by deleting some unnecessary variables and their coefficients in the process of estimation. The present thesis derives an algorithm for implementing Lasso to SEM and carry out several simulation studies to evaluate its performance, comparing the result with that of existing estimation methods such as maximum likelihood principle and Bayesian Lasso SEM.

As the evaluation criteria, we observe several discrepancy indices including OD, DA, and DE. Though these indices are not observable in practice, we can

approach these indices by means of simulation studies including generation of population data. Furthermore, we investigate whether Lasso is able to reduce DE, which indicates the sampling variability in model fitting, and OD, which represents the generalizability/reproducibility of the model.

Additionally, we also compute SD and a variety of goodness of fit indices, and conduct a correlation analysis so as to figure out which of them have close relationships with discrepancies above, especially OD and MSE as generalizability indices. If there is one, the index can be exploited in practice as an indirect criterion indicating the reproducibility of the model, by its connection with OD or MSE.

Thesis Organization

In this introduction, we discussed the issue of reproducibility in psychological researches and the need for the analytical/statistical approach on this issue.

In **Chapter 1**, we introduce the structural equation modeling with its estimation, criteria for model evaluation, and the generalizability/reproducibility issue relevant to the model.

Chapter 2 discusses the philosophy and several methods of regularization. Also its connection with Bayesian approach is covered. We end the chapter emphasizing the need for implementing L_1 regularization(Lasso) to SEM and suggesting several algorithms that can be exploited to accomplish our purpose.

Chapter 3 provides a brief introduction to BLasso SEM, which is an abbreviation for Bayesian Lasso SEM, as one of preceding research results. Important limitations of the method are suggested.

In **Chapter 4**, we derive an algorithm for regularizing SEM via the Lasso. Since SEM is the combination of two different types of models, we propose 'Double EM-algorithm' to estimate both models simultaneously. Several further issues in fitting L_1-regularized SEM are also discussed.

In **Chapter 5**, methods and conditions for simulation studies are provided. This includes the population-generating process, parameter values and model specifications, research conditions including sample sizes and magnitudes of one of parameter matrices, and various fit indices investigated in the studies.

Chapter 6 provides the result of simulation studies.

Chapter 7 summarizes the present thesis, and also gives its limitations, implications, and suggestions for future research.

Chapter 1

Structural Equation Modeling

1.1 Introduction to SEM

The Structural Equation Modeling(SEM), or Covariance Structure Modeling(CSM) consists of two models. The one, which is called 'Measurement Model', deals with the causal relationship between psychological constructs and their measurements. The other contains the relationship among the latent variables and usually is named 'Structural Model'.

1.1.1 Measurement Model Part

The measurement model in SEM is described as follows.

$$Y = \Lambda\Omega + \epsilon \tag{1.1.1}$$

Here, Y is a $p \times n$ matrix whose columns are y_i's, the $p \times 1$ response vectors for subjects. Each row in Y contains an individual variable, which are called 'Measurement Variable'. Sometimes this is also referred to as 'Observed Variable' or 'Manifest Variable'. Λ is a $p \times q$ matrix that contains path coefficients or regression coefficients. Ω is a $q \times n$ matrix whose columns are ω_i's, the $p \times 1$ latent scores for subjects and ϵ is a $p \times n$ error term matrix.

$$\begin{aligned}
\mathbf{Y} &= \begin{pmatrix} y_1, & y_2, & \ldots, & y_n \end{pmatrix} \\
\mathbf{\Omega} &= \begin{pmatrix} \omega_1, & \omega_2, & \ldots, & \omega_n \end{pmatrix} \\
\epsilon &= \begin{pmatrix} \epsilon_1, & \epsilon_2, & \ldots, & \epsilon_n \end{pmatrix}
\end{aligned} \tag{1.1.2}$$

For i-th subject, this model can be described as follows.

$$\begin{aligned}
y_i &= \Lambda\omega_i + \epsilon_i \\
i &= 1, \ldots, n
\end{aligned} \tag{1.1.3}$$

And in this equation, we assume

$$\begin{aligned}
y_i &\sim N_p(0,\ \Lambda\Sigma_\omega\Lambda^T + \Psi_\epsilon) \\
\omega_i &\sim N_q(0,\ \Sigma_w) \\
y_i|\omega_i &\sim N_p(\Lambda\omega_i,\ \Psi_\epsilon)
\end{aligned} \tag{1.1.4}$$

This model consist of p equations, which define the relationship between p observed response variables and corresponding latent variables. The response variable y is often called an 'observed variable', a 'response variable', or a 'manifest variable'. And the latent score ω is also called as a 'factor score'.

This model resembles the linear regression model, but there is a big difference between them. That is, in measurement model, the regressors(or independent variables) are not observable. Commonly, in psychology, the latent variable ω represents 'Psychological Construct' such as depression, anger, self-esteem and so on. Since we are not able to observe these constructs, we develop and use some measurement instruments like questionnaire, response times and other measures. The measurement model help us to analyze our measurement instruments to investigate whether these tools measure the target latent variable well, or to conduct various statistical inferences on the factors.

This model is widely known as the 'Factor Analysis'(FA) model. In the FA model, usually we use ξ_i and Φ, instead of ω_i and Σ_ω respectively. And usually Λ is called a 'factor loading' matrix.

The following figure represents a typical two factor model with each latent variable having four measurement variables.

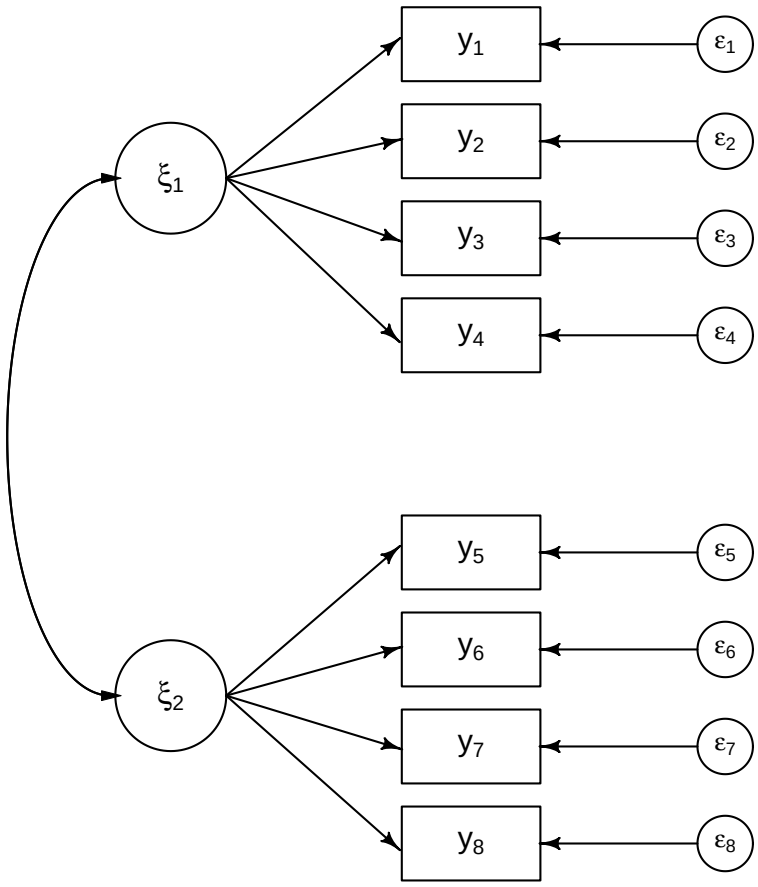

Figure 1.1: An Example of Factor Analysis Model

1.1.2 Structural Model Part

The structural model in SEM represents relationships among the latent variables. These variables can be divided into the following two discrete categories. The first one is the latent 'Endogenous' variable, which is affected by other variables in the model. These variables are responses in the structural model. So, they act as dependent variables in their respective regression equations. But they also can be used as regressor for other endogenous variables. This kind of variable is usually denoted as η

The other one is the latent 'Exogenous' variable, which is not affected from other variables. Thus each of these variables is determined from the outside of the model, and has a linear influence on some of endogenous variables. Therefore, exogenous latent variables are included in the model only as regressors. We denote these variables as ξ.

One of most well known SEM model, LISREL, defines the structural model as follows.

$$\eta = B\eta + \Gamma\xi + \zeta \tag{1.1.5}$$

Here, η is a $q_1 \times n$ matrix whose column is η_i, a $q_1 \times 1$ latent endogenous variable and ξ is a $q_2 \times n$ matrix whose column is ξ_i, a $q_2 \times 1$ latent exogenous variable. Therefore, we have $q = q_1 + q_2$ latent variables in this model. B is a $q_1 \times q_1$ matrix representing the relationships among the elements of η_i, and Γ is a $q_1 \times q_2$ matrix representing the effect of latent exogenous variables on latent endogenous variables. A $q_1 \times n$ matrix ζ includes ζ_i's, the $q_1 \times 1$ error terms in the structural model.

Using these notations, the structural model for i-th response is

$$\eta_i = B\eta_i + \Gamma\xi_i + \zeta_i, \quad i = 1, \dots n. \tag{1.1.6}$$

Here, some distributional assumptions are added.

$$\begin{aligned} \xi_i &\sim N_{q_2}(0, \Phi) \\ \zeta_i &\sim N_{q_1}(0, \Psi_\zeta) \\ \text{where } \Psi_\zeta &= diag(\psi_{\zeta_k})_{k=1,\ldots,q_1} \end{aligned} \tag{1.1.7}$$

To clarify the connection between the measurement model and the structural model, (1.1.6) can be described in the slightly different way.

$$\begin{aligned} \omega_i = \begin{pmatrix} \eta_i \\ \xi_i \end{pmatrix} &= \begin{pmatrix} B & \Gamma \\ 0 & I \end{pmatrix} \begin{pmatrix} \eta_i \\ \xi_i \end{pmatrix} + \begin{pmatrix} \zeta_i \\ 0 \end{pmatrix} \\ &= \Lambda_\eta \quad \omega_i + \zeta_{\omega i} \end{aligned} \tag{1.1.8}$$

As mentioned in the measurement model,

$$\omega_i \sim N_q(0,\ \Sigma_w) \tag{1.1.9}$$

where

$$\begin{aligned} \mathbf{\Sigma}_\omega &= \begin{pmatrix} \Sigma_{\eta\eta} & \Sigma_{\eta\xi} \\ \Sigma_{\xi\eta} & \Sigma_{\xi\xi} \end{pmatrix} \\ &= \begin{pmatrix} (I-B)^{-1}(\Gamma\Phi\Gamma + \Psi_\zeta)((I-B)^{-1})^T & (I-B)^{-1}\Gamma\Phi \\ \Phi\Gamma^T((I-B)^{-1})^T & \Phi \end{pmatrix} \end{aligned} \tag{1.1.10}$$

Derivation of this matrix is put off to the Appendix C. The figure in the next page represents an example of a structural equation model. This is quoted from Browne and MacCallum(2004), with some modifications for easy understanding. The model has three latent variables, two of those are endogenous(η) and remaining one is exogenous(ξ). And there are equally two measurement

variables per each latent. Also note that each of endogenous latent variables has its own error term(ζ).

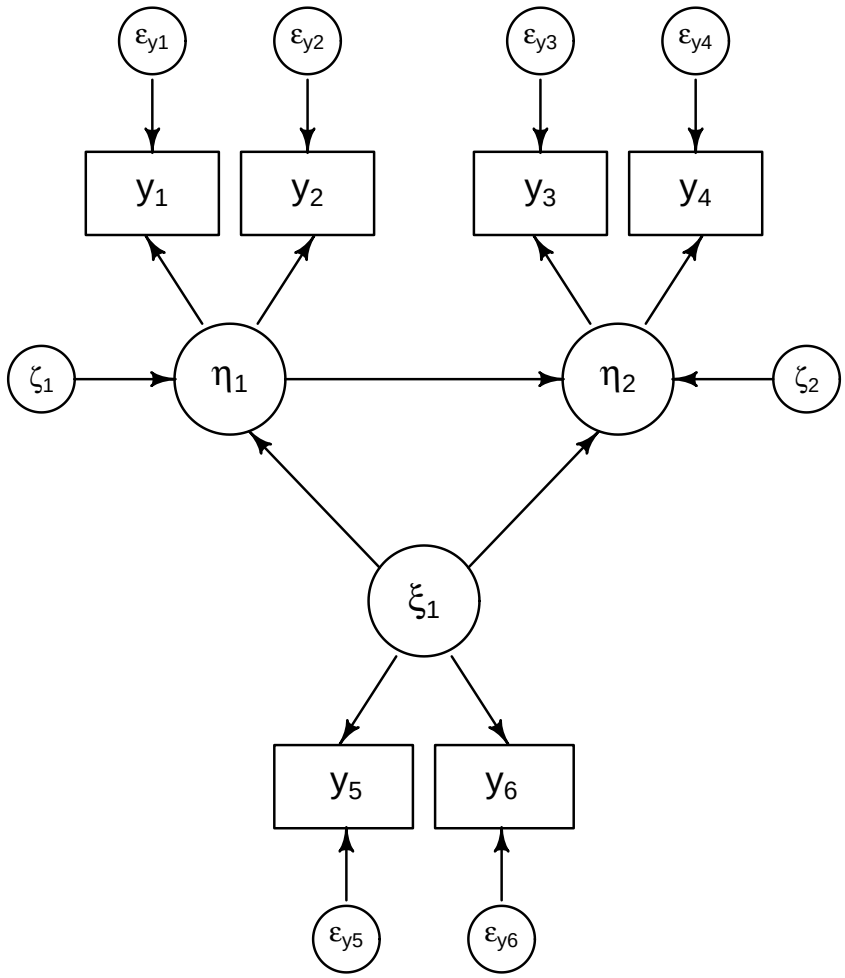

Figure 1.2: An Example of Structural Equation Model

It is obvious that both models in the SEM have the same mathematical formula as in the linear regression model. Since the SEM can deal with more complex structure than a typical regression analysis, this usually is understood as an extension or a generalization of the linear regression models. However, there exist several fundamental differences between two class of models. The first one is, as we described in the measurement model part, the SEM contains latent variables. Measurement model has observed responses, but its regressors are latent. Even worse, structural model consists of only the latent variables. This makes the practical analysis and estimation of the model difficult and challenging. However, this enables us to deal with the hypothetical network among the unobservable variables such as psychological constructs.

SEM is different with linear regression analysis in that it is a complex system of a large number of equations. In linear regression analysis, we usually have interests on the one equation that has lots of regressors for only one response variable. However, SEM deals with a string of equations that commonly have a few independent variables. Furthermore SEM concerns the case when some variables are affected by and also affect others. In most cases, latent endogenous variables have this kind of sequential relationships as they are used as responses in some equations and also as regressors in other equations. Hence the variables in SEM are complicatedly entangled, unlike the multiple regression where all the regressors have linear relationship with their response.

With the above points of distinction, SEM gives us a capability of analyzing more complex structure among the interested variables. Also this model makes it possible for us to investigate into the underlying relationships consisting of latent variables.

1.2 Estimation of SEM

As we described in the previous section, SEM contains a number of parameters. Therefore its estimation problem can easily be over-complex. The most widely accepted method is to use 'Discrepancy Function'. In SEM, estimated parameter values can be used to derive the 'Implied Covariance Matrix' which contains the expected relationship or 'Covariance/Correlation Structure' among the interested variables. This matrix can be computed with the following formula.

$$\Sigma = \Lambda \Sigma_\omega \Lambda^T + \Psi_\epsilon \tag{1.2.1}$$

The implied covariance matrix is calculated with estimates of Λ, Ψ_ϵ, and Σ_ω are substituted in the above formula. Also, Σ_ω in the formula can be obtained by equation (1.1.10). In factor analysis model, Σ_ω is replaced with Φ.

So-called 'good' estimation must be able to recover and reconstruct the true relationship at the level of population into the implied covariance matrix. Hence, some functions, denoted as F, are defined to measure the distance, or 'Discrepancy' between the implied covariance matrix $\hat{\Sigma} = \Sigma(\hat{\theta})$ and the covariance matrix of population Σ_0. Usually the value of this function is denoted as $F(\Sigma_0, \hat{\Sigma})$. Based on this term, the estimates of θ that minimize F-vale will be adopted as good estimates.

However, in real-world experiments where the whole population is not available, scientists are not able to make use of Σ_0. Instead, sample data can be considered as the estimates of population they are sampled from(Browne & MacCallum, 2004). In fact, it is well known that the sample and its nonparametric distribution function are the nonparametric maximum likelihood estimates of the population and its distribution function, respectively(Wasserman, 2006).

Based on this theoretical result, $F(S, \hat{\Sigma})$, the discrepancy between the sample covariance matrix S and $\hat{\Sigma}$ can be viewed as an estimates of $F(\Sigma_0, \hat{\Sigma})$, and is supposed to be minimized so that corresponding estimates $\hat{\theta}$ are produced.

There are a couple of conditions that the discrepancy function F should satisfy(Browne, 1984; Bollen, 1989).

i) $F(S, \Sigma) \geq 0$

ii) $F(S, \Sigma) = 0$ if and only if $\Sigma = S$

iii) $F(S, \Sigma)$ is continuous in S and Σ.

It is proven that estimation methods by minimizing this kind of discrepancy function can produce the consistent estimators of θ(Browne, 1984; Steiger, Shapiro, & Browne, 1985; Bollen, 1989).

The following is a list of the most commonly used types of the discrepancy functions(Bollen, 1989; Browne et al., 2002; Browne MacCallum, 2004).

OLS Discrepancy Function

$$F_{OLS}(S, \Sigma(\theta)) = \frac{1}{2} tr((S - \Sigma(\theta))^2) \tag{1.2.2}$$

This function can be understood as a matrix-version of the 'Ordinary Least Squares(OLS)' loss function, or the sum of squares of residuals in the linear regression model. The great advantage of OLS discrepancy function is its simplicity and robustness. However, since it has no distributional assumptions, any statistical inferences including computation of standard errors and confidence intervals, and the hypothesis testing are unavailable(Browne MacCallum, 2004).

ML Discrepancy Function

$$F_{ML}(S, \Sigma(\theta)) = \log(|\Sigma(\theta)|) - \log(|S|) + tr(S\Sigma(\theta)^{-1}) - p \tag{1.2.3}$$

'Maximum Likelihood Estimation'(MLE) is one of the most dominant methods for statistical estimation and inference over the past few decades. If some distributional assumptions can be made on the distribution of our data, its likelihood function can be derived easily, and used as a ML-type loss function for the estimation of the interested parameters.

SEM usually requires the sample data to follow the Normal distribution. Or the Wishart distribution of its covariance matrix is also sufficient, which is the reason why some texts on SEM state the ML methods as 'Maximum Wishart Likelihood'(MWL) method.

The function (1.2.3) is considerably connected to the likelihood function of SEM. In fact,

$$\begin{aligned}(n-1)F_{ML} \simeq -2LL &= -2\log(\frac{L_1}{L_2}) \\ &= -2(\log(L_1) - \log(L_2))\end{aligned} \tag{1.2.4}$$

where L_1 is a likelihood function of the current model, and L_2 indicates that of the saturated model which satisfies $\hat{\Sigma} = S$. The most attracting benefit of using ML discrepancy function comes out from this relationship. That is, this method is highly relevant to general theories on maximum likelihood-based estimation and testing which are proven and organized well in statistics. The term L_1/L_2 is called 'Likelihood Ratio'(LR), and the statistical testing that uses $-2LL$ as its test statistic is called 'Maximum Likelihood Ratio Test'(MLRT). It is proven that this test statistic asymptotically follows χ^2 distribution(Kim, 2012). Therefore,

$$(n-1)F_{ML} \rightsquigarrow \chi^2(df) \tag{1.2.5}$$

Hence, from this distribution and relational theorems, the estimation of SEM using the ML discrepancy function is able to take advantages of many great statistical properties of the general ML method.

GLS Discrepancy Function

$$F_{GLS}(S, \Sigma(\theta)) = \frac{1}{2} tr(W^{-1}(S - \Sigma(\theta))^2) \tag{1.2.6}$$

Generalized Least Squares(GLS) is a generalized version of OLS, which is defined as giving weights to the squares of residuals. Commonly, the unbiased sample covariance matrix S is used as weight matrix W. Sometimes $\hat{\Sigma}_{ML}$, the implied covariance matrix in ML method is also used for the weight (Bollen, 1989; Browne et al., 2002).

It is well known that, like F_{ML}, F_{GLS} is also related to the asymptotic distribution as follows(Bollen, 1989).

$$(n-1)F_{GLS} \rightsquigarrow \chi^2(df) \tag{1.2.7}$$

This enables GLS method to enjoy the great statistical properties like ML method. However, since there are a little benefits in GLS over ML, this method is barely used, especially in small samples(Browne & MacCallum, 2004). Only advantage of GLS is that it is slightly more robust than ML. For example, when the sample covariance S is not positive-definite, GLS method works normally while ML method fails.

1.3 Fit Indices for Model Evaluation

Since SEM contains a large number of parameters, its main interest is in evaluating the degree of how well the current model explains the interested phenomenon, rather than each of parameters. For this purpose, the classic method of hypothesis testing was suggested originally. This uses the asymptotic distribution of discrepancy functions as described in (1.2.5) and (1.2.7). The null and alternative hypotheses is as follows.

$$H_0 : \Sigma = \Sigma_\theta \qquad \text{vs} \qquad H_1 : \Sigma = S.$$

Here H_0 means that the current model perfectly explains the true covariance structure in the population while H_1 means that there is no systematic covariance structure in the interested phenomenon. From the meaning in the null hypothesis, this testing method is usually called the 'test of perfect fit'.

However, in this hypothesis testing, the hypothesis which contains researcher's model so that they want to shore up, is H_0 not H_1 as in the general null hypothesis testing method. In this method, researchers' model is accepted when H_0 is not rejected. Actually, the evidence we can find in this testing is that the current model is not too inaccurate to be rejected, not the one that says the current model is excellent enough to explain the interested phenomenon and support the researchers' theory. It is well known that this hypothesis testing rejects H_0 too easily since it is unrealistic that the current model has perfect explanations on the real-world covariance system. Test of perfect fit also rejects H_0 easily as long as the sample size is quite small. This is due to the fact that p-value, a statistical criterion of accepting or rejecting the hypothesis, is a function of the sample size. For all those reasons, it is recommended not to use this hypothesis testing method.

As an alternative, a lot of model fit indices were suggested by the psychometricians. These indices can be used to evaluate SEM models in the manner that if a value of the index is larger or smaller than its criterion, then the model is accepted as excellent enough, and otherwise the model is defeated. Some indices evaluate the model in the different way. An index of this kind is used in comparing several candidate models in that the one which yields the smallest value of the index is adopted.

The following is the list of widely used fit indices with an appropriate categorization. Note that the mathematical definition of each indices are gathered in table 1.1 below.

1) Absolute Fit Indices

Hu and Bentler(1995) explains that absolute fit indices "directly assess how well an a priori model reproduces the sample data." According to this explanation, absolute fit indices can be understood to be designed to measure how well the currently estimated model describes the covariance structure in the sample data. Among this type of indices are RMSEA(Root Mean Squared Error of Approximation), SRMR(Standardized Root Mean squared Residual), Mc(McDonald's Centrality Index), $\hat{\gamma}$, etc.

Perhaps the most famous one is RMSEA, which is known for the fact that it has a distribution related to χ^2 distribution of $(n-1)F_{ML}$ or $(n-1)F_{GLS}$. This distribution enables us to use the 'test of close fit', which is an alternative to the test of perfect fit(Browne & Cudeck, 1992). This testing method uses RMSEA as its test statistics, and its hypotheses are as follows.

$$H_0 : RMSEA \geq .05 \qquad \text{vs} \qquad H_1 : \text{not} \quad H_0.$$

Even though this testing also has the problem that researchers' hypothesis is H_0, not H_1, it is considered as being quite better than the test of perfect fit. However, RMSEA is better known for its simple criteria of assessing goodness of model fit(Browne & Cudeck, 1992).

[, .05] : Close Fit
[.05, .08] : Reasonable Fit
[.08, .10] : Mediocre Fit
[.10,] : Unacceptable

Another strength of RMSEA is that it has a confidence interval, so that researchers can make a more stable decision on goodness of their model(Browne & Cudeck, 1992).

SRMR is also widely used when evaluating SEM models. In fact, in their simulation study, Hu and Bentler(1998) reveals that SRMR performs best in a variety of situations and it was the most sensitive to the case when the model misspecifies factor covariances. Additionally, since it is simply based on the sum of squared residuals, there is no assumption for SRMR to perform best. Therefore this index can be used in any situation with less concern on potential malfunctioning.

2) Incremental Fit Indices

On the contrary to absolute fit indices which assess the current model without any reference model, incremental fit indices are designed to evaluate our model in comparison with a baseline model. This indices "measures the proportionate improvement in fit by comparing a target model with a more restricted, nested baseline model", described in Hu and Bentler(1998). In their explanation, 'nested' indicates the case that the parameter space of one model

is a subset of that of the other model. Only the model satisfying this condition can be used as a reference model when we assess the current model. In SEM, most choice of the baseline model is the null model $\Sigma = diag(\sigma^2)_{i=1,\ldots,p}$, which implies that there is no covariance among the interested variables. Accordingly most of incremental fit indices are defined using terms from the current model and those from the null model.

Among this type of indices are NFI(Normed Fit Index), TLI or NNFI(Tucker-Lewis Index, or Non-Normed Fit Index), CFI(Comparative Fit Index), RFI(Relative Fit Index, or BL86), IFI(Incremental Fit Index, or BL89), RNI(Relative Noncentrality Index), and so on.

3) Information Criteria : AIC and BIC

The next two indices, AIC and BIC, were developed in the field of information theory. Akaike developed 'Akaike Information Criterion'(AIC) while extending the maximum likelihood principle to model identification(Akaike, 1973; Akaike, 1974; deLeeuw, 1992; Akaike, 1987). His result can be regarded as an estimate for expected probabilistic Kullback-Liebler divergence(or negentrophy). Thus, if we choose the model that has the smallest value of AIC, this lead us to the model with the smallest Kullback-Leibler divergence. This is also referred to as an estimate of statistical risk(Wasserman, 1974) or a generalization of expected prediction error(Akaike, 1974). Since this index can be computed instantly in a variety of statistical methods which are based on the concept of likelihood, including multiple regression, factor analysis model, principle component analysis, analysis of variance, and time series analysis(Akaike, 1973; Akaike, 1987), this has been widely used by researchers from the time it is proposed.

Even though it might seem algebraically similar with AIC, 'Bayesian In-

formation Criterion', or BIC, was developed by almost completely different derivation, which is related to the context of Bayesian estimation. That is, as its developer Schwarz revealed, this criterion was founded as the result of "studying the asymptotic behavior of Bayes estimators under a special class of priors"(Schwarz, 1978).

Among a number of candidate models, we can choose the one that makes AIC or BIC minimized as our selection. Hence, with these criteria, model estimation and selection can be done based on the simple scalar-valued index. Since these indices take the trade-off between the model fit and model complexity into consideration, they are also usually regarded as indices for generalizability of the model. Another valuable strength of this kind of criterion is that it is easily computable when we can calculate the likelihood of interested models.

4) Expected Cross Validation Index : ECVI

Cross-validation methods have been widely used in a variety of fields for obtaining a more generalizable model that performs better in prediction problem. Since it was proposed first for the linear regression analysis, it can be extended to SEM directly. The procedure was applied first to SEM by Cudeck and Browne(1983) and an index they called 'cross-validation coefficient' was studied. The most challenging problem in carrying out cross validation is that it needs the current sample split into calibration(training) sample and validation(test) sample. The former is used in fitting models, and the latter is used in evaluating them. The cross validation coefficient is calculated as a value of discrepancy function with an implied covariance matrix from the calibration sample as a target, and a sample covariance matrix from the validation sample as an input. More on general cross validation methods will be illustrated in

Chapter 4.

Dividing a sample into two distinct sets causes an inefficiency in estimation and testing due to the reduced sample size. In order to avoid this side effect, ECVI, an abbreviation for Expected Cross Validation Index, was suggested as a single sample cross validation coefficient(Browne & Cudeck, 1989; Browne & Cudeck, 1992; Browne, 2000). As its name implies, this index is the expectation of cross validation coefficient over both of calibration and validation samples. The estimator of ECVI, c, is defined as a simple sum of observed discrepancy and an estimator of difference between the cross validation coefficient and calibration sample discrepancy.

$$c = F(S, \hat{\Sigma}) + 2q/(n-1) \tag{1.3.1}$$

Note that this is computable without any validation sample. Meanwhile, It is well-known that in linear regression analysis, the prediction validity based on log-density assessment is asymptotically equivalent to AIC under weak conditions, when it is computed by leave-one-out cross validation. The same relation is validated in SEM; That is, ECVI can also be understood as a rescaled version of AIC, when the maximum likelihood type discrepancy function (1.2.3) is exploited.

$$\begin{aligned} AIC &= n \times c - 2\log(L_2) \\ c &= \frac{1}{n}(AIC + 2\log(L_2)) \end{aligned} \tag{1.3.2}$$

These equations are followed from equations (1.2.3) $\sim$ (1.2.4), (1.3.1), and definitions of AIC and ECVI in table 1.1. The relationship between two indices is reasonable since AIC is defined as a measure of generalizability and ECVI is the outcome of studies on cross-validation for choosing a more generalizable

model. In conclusion, both indices can be criteria for decision as to whether the model will produce stable results when it fit to the distinct samples.

The following table in the next page contains model fit indices described above. Here, subindex 'null' indicates the null model described in the explanation of incremental fit indices. And subindex 'saturated' indicates the saturated model $\Sigma = S$, which implies that there are some covariational relationships among the interested variables, but it is impossible to explain it systematically. Thus no hypothetical covariance structure is appropriate to explain data, but sample covariance itself is.

Also note that χ^2 indicates the value of $(n-1)\hat{F} = -2LL$ that follows asymptotic χ^2 distribution when we use ML or GLS discrepancy function as explained. And df denotes the 'degree of freedom' of this χ^2 distribution and is obtained by subtracting the number of free parameters t^* from the number of distinct elements in the sample covariance matrix S, $p^* = p(p+1)/2$. At last, L indicates the likelihood of the model.

Name	Definition	References
Absolute Fit Indices		
RMSEA	$\sqrt{\frac{\max(\hat{F} - \frac{df}{n-1},\ 0)}{df}}$	Steiger & Lind(1980) Hu & Bentler(1998)
SRMR	$\sqrt{\sum_{i=1}^{p}\sum_{j=1}^{i}\left[(\frac{s_{ij}-\hat{\sigma}_{ij}}{s_{ii}s_{jj}})^2\right] / \left(\frac{p(p+1)}{2}\right)}$	Bentler Hu & Bentler(1998)
Mc	$exp(-\frac{1}{2}\left[\frac{\chi^2_{current} - df_{current}}{n-1}\right])$	McDonald(1989) Hu & Bentler(1998)
$\hat{\gamma}$	$p/[p+2\left[\frac{\chi^2_{current} - df_{current}}{N-1}\right]]$	Hu & Bentler(1998)
Incremental Fit Indices		
NFI	$\frac{\chi^2_{null} - \chi^2_{current}}{\chi_{null^2}}$	Bollen(1989) Jöreskog & Sörbom(1993) Hu & Bentler(1998)
TLI(or NNFI)	$\frac{\chi^2_{null}/df_{null} - \chi^2_{current}/df_{current}}{\chi^2_{null}/df_{null} - 1}$	Bollen(1989) Jöreskog & Sörbom(1993) Hu & Bentler(1998)
CFI	$1 - \frac{max(\chi^2_{current} - df_{current})}{max(\chi^2_{current} - df_{current},\ \chi^2_{null} - df_{null}, 0)}$	Bollen(1989) Jöreskog & Sörbom(1993) Hu & Bentler(1998)
RFI(or BL86))	$\frac{\chi^2_{null}/df_{null} - \chi^2_{current}/df_{current}}{\chi^2_{null}/df_{null}}$	Bollen(1989) Jöreskog & Sörbom(1993) Hu & Bentler(1998)
IFI(or BL89)	$\frac{\chi^2_{null} - \chi^2_{current}}{\chi^2_{null} - df_{current}}$	Bollen(1989) Jöreskog & Sörbom(1993) Hu & Bentler(1998)
RNI	$\frac{(\chi^2_{null} - df_{null}) - (\chi^2_{current} - df_{current})}{\chi^2_{null} - df_{null}}$	Bollen(1989) Jöreskog & Sörbom(1993) Hu & Bentler(1998)
Information Criteria		
AIC	$-2log(L) + 2\,t^*$	Akaike(1973, 1987)
BIC	$-2log(L) + t^* log(n)$	Schwarz(1978)
Expected - Cross Validation Index		
ECVI	$\hat{F} + \frac{2t^*}{n-1}$	Browne & Cudeck(1992) Browne(2000)

Table 1.1: List of Fit Indices

1.4 Reproducibility and Generalizability Issues in SEM

Thus far, fundamentals on SEM, including mathematical description of the model, methods of estimation, and indices for evaluation of the model were illustrated. In this section, we shall discuss what is a good model and how we achieve it at least approximately, especially in the context of psychometric modeling and SEM. This issue is highly related to the model selection problem in statistics – Among the large number of candidate models, which one we choose?

Consider the ultimate purpose of inferential statistics. It attempts to find the best way to comprehend the underlying process in which we are interested, under the limited sample size. Thus the problem is to achieve an appropriate explanation of characteristics inherent in the current sample and figure out those are generalizable to the population and the real-world phenomenon.

In this regard, there have been several suggestions on the virtues of a good model. At first, from the common-sense point of view, the model should fit the current sample quite well. This is a topic of goodness of fit, which is a prerequisite of the model to explain well the phenomenon. With a model that fits poorly to the sample, we have almost nothing to talk about the underlying process generating our population.

However, model fit cannot be an absolute criterion for the decision. Note that many of the goodness of fit indices are improved simply by incorporating additional variables or parameters. Thus more complex models tend to fit the current sample better than simpler ones. However, many previous researchers pointed out that these models with a large number of parameter suffer from their excessive variability(Myung & Pitt, 1997; Myung, 2000; Bishop, 2006; Hastie, Tibshirani & Friedman, 2008). Models with overparameteriza-

tion may explain the current data well, but the result is unlikely to be replicated when it is fitted to other independent samples. This claim seems valid and reasonable considering a natural behaviors of parameters in a statistical model. When only a few parameters are included in the model, they capture main trends in the current sample to establish a well-fitted result. Plus, sometimes Researchers try to add more parameters, some of which may turn out to be superfluous, to obtain a better fit. In this case, the latest additions end in capturing minor errors and randomness in the sample when the model becomes out of proportion to the sample size in terms of its complexity. Since these minor thing only belong to the current data, the added parameters perform poorly when the same model is fitted to another distinct sample.

Furthermore, complex models may not be useful in a practical aspect since it is difficult to interpret them. When dealing with models with a small number of parameters, researchers are able to come to a clear understanding on the relationship among all the dependent and independent variables engaged in the present model. Thus these simple models can be effective tools for researchers when they try to make parsimonious and straightforward explanation on phenomena they are interested in.

However, when it comes to the case they are dealing with a quite complex model with a huge number of parameters, the variables involved in the model are likely to form an awfully complicated relation structure. Maybe this model has the best fit, we cannot extract anything valuable in practice from this result.

To conclude, criteria for attaining a 'good' model should encompass not only a conventional goodness of fit concept, but also generalizability and interpretability which can be achieved from simple and parsimonious models. However, the latter two concepts are not clearly measurable characteristics. This causes hardship for a discussion on the model selection issue. Nonethe-

less there are several frameworks for model evaluation, which make it feasible to conceptualize the issue and related characteristics.

Among them, Cudeck and henly(1991)'s contributions are noteworthy with regard to the structural equation modeling. The following figure shows the framework they suggested for representing several covariance matrices we should bear in mind when analyzing SEM. Note that there are four important covariance matrices in SEM. The first is the population covariance matrix Σ_0. In SEM we assume that data is from the multivariate normal distribution with covariance matrix Σ_0 or sample covariance matrix S is from the wishart distribution with parameter matrix Σ_0. Main purpose of SEM is to investigate the structure inherent in this matrix.

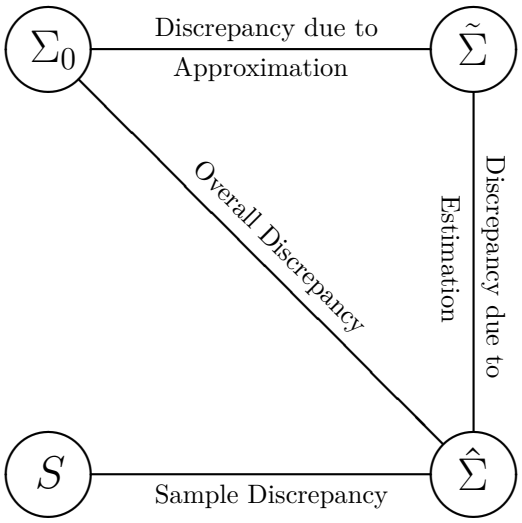

Figure 1.3: Types of Discrepancy

Cudeck and Henly(1991) considers three conceptualization of the structure (Cudeck & Henly, 1991; Cudeck & Browne, 1992).

a) $\Sigma_0 =$ arbitrary covariance matrix

b) $\Sigma_0 = \Sigma_0(\theta_0)$

c) $\Sigma_0^* = \Sigma_0(\theta_0) + E$

Structure a) implies that relationships among the variable are unstructured. This assumes that the population are generated from completely unknown operation. Structure b) indicates that there exists an 'Operating Model' $\Sigma_0(\cdot)$ that is able to reproduce Σ_0 if we know the value of θ_0. By definition, this model can be called as 'True Model' or 'Data generating process' in Preacher et al.(2013). In this case, Σ_0 can be explained exactly by some systematic structure. But note that it is implausible to say we discover this structure perfectly, since the population process is know to be highly overcomplex and infinite-dimensional. The last conceptualization, c), reflects the case that there is a systematic structure that affects the generation of Σ_0^*, but there also exist a huge number of minor factors that give contribution to Σ_0^*. The error occurred from these factors is called 'Model Error' and is represented by the matrix E in equation c). In this case it is impossible to reproduce Σ_0^*. All we can do is trying to approximate the structure embedded in Σ_0^*. The term 'Operating Model' can also be used in this case, but it has slightly different meaning of 'Quasi-true Model' or 'Quasi-true Data Generating Process'. These terms involve the idea that the model cannot regenerate the population covariance matrix, but it is the best approximating model of the true mechanism.

In practice, we set up one or more models $\Sigma_k(\cdot)$ $(k = 1, \ldots, m)$ in order to investigate Σ_0(or Σ_0^* in case we accept the concept c)). Each model may have

different parameters so that it implies different hypothetical structures. Using these models, SEM produces two matrices.

The one is $\tilde{\Sigma} = \Sigma_k(\theta_0)$. This is the matrix regenerated by our model with unknown parameter θ_0. Thus this is our potentially best approximation to population covariance matrix Σ_0. However, this matrix is also unavailable since we don't know about θ_0 in practice.

The sample level counterpart of $\tilde{\Sigma}$ is the implied covariance matrix $\hat{\Sigma} = \Sigma_k(\hat{\theta})$ described before in Section 1.2. This matrix is reproduced by means of our model as minimizing the value of discrepancy between sample covariance S which is the fourth matrix in our consideration, and $\hat{\Sigma}$ itself. The implied covariance matrix in our best approximation which is available in practice. Hence, the main purpose of SEM can be said to be obtain the model $\Sigma_k(\cdot)$ able to reproduce $\hat{\Sigma}$ such that best approximates the covariance structure in Σ_0.

Figure 1.3 illustrates well the relationship among those matrices. The upper side of the figure represents the population level where matrices belonging to here is not observable. They are available only when we know θ_0 and $\Sigma_0(\cdot)$. In case Σ_0 follows a) or c), we can achieve this matrix even though this condition are satisfied. The bottom side shows the sample level where the practical estimation procedure is involved. Both S and $\hat{\Sigma}$ are computable with our sample.

Furthermore, as the figure reveals, there are several discrepancy concepts between these matrices. At first, 'Discrepancy due to Approximation(DA)' is defined as a discrepancy between Σ_0 and $\tilde{\Sigma}$. As described before, regenerating Σ_0 with our model is not possible. Hence there exists a dissimilarity between two matrices. This concepts is assumed to represent the plausibility(Browne & Cudeck, 1992), or the verisimilitude(Preacher et al., 2013) of the current model. Note that in case we correctly specify our model, $\Sigma_k(\cdot) = \Sigma_0(\cdot)$, DA is equal to discrepancy resulting from the model error. Conceptually this is close to 'bias'

of the estimator in statistical inference. Also note that this discrepancy does not depend on samples.

'Discrepancy due to Estimation(DE)' refers to the difference between $\tilde{\Sigma}$ and $\hat{\Sigma}$. This implies the variability occurred due to sampling procedure, not to a lack of accuracy of the model. This concept may reveal the reader 'variance' of the estimator.

'Overall Discrepancy(OD)' is the discrepancy between Σ_0 and $\hat{\Sigma}$. Thus this indicates the quality of our model estimation based on the current sample. Moreover, in the figure it seems that OD path is related to the path connecting DA and DE. In fact, it is proven that $OD = DA + DE + o(N^{-1})$(Browne & Cudeck, 1992; Browne, 2000; Preacher et al., 2013). This formula makes us recall the concept of 'bias-variance decomposition' or 'bias-variance trade-off' in statistics, which says that 'Mean Squared Error(MSE)', an index indicating the excellence of an estimator, is decomposed into bias squared plus variance, and when the one is made to be decreased, the other tends to increase. In fact, OD and MSE are closely related in that hey are regarded as indices of generalizability of the model and its result. This point will be discussed later in this section for OD, and in Section 2.1 for MSE.

The last discrepancy concept is 'Sample Discrepancy(SD)', which indicates the deviance between S and $\hat{\Sigma}$. This represents a goodness of fit of our model in the current sample. Notice that SD is the only observable discrepancy in practical analysis among those four discrepancy concepts.

It should be noted that these discrepancy concepts play important roles in estimation and evaluation of SEM models. Especially, they can give us the direction to achieve a good model we described above. For estimation, SD is almost all of the model fitting procedure, since this is the only discrepancy we are able to observe. Remind that implied covariance matrix $\hat{\Sigma}$ and relate

parameter estimates are attained by minimizing SD. Since many of fit indices described in Section 1.3. is built on the relationship to SD, many researchers devote their efforts to reduce SD. Careful reflection and meticulous study on how to build a model can be helpful. But hits can be achieved simply by adding more parameters and variables to their model, some of which are unnecessary and have no theoretical backgrounds.

However, SD is not a criteria for a good model. It is not directly related to our objective matrix Σ_0. Thus decreasing SD by inappropriate methods never guarantees our goal. Since SD is a biased estimator of DA(McDonald, 1989; Browne & Cudeck, 1992), these efforts may be able to reduce DA. In the population level, where the population is known to be infinite-dimensional and contains not only the main process but also a number of randomness, it seems that augmenting the model with superfluous factors and variables can be of help achieving the true model. However, in the sample level, this lead us to construct overly complex model which the current sample size lack the capability of fitting it. This in turn aggravates stability of the model and inflates DE severely. Approximated formula of expected DE, $E(F(\tilde{\Sigma}, \hat{\Sigma})) = \frac{q}{n}$(Browne & Cudeck, 1992; Browne, 2000), where q is the number of parameters here, clarifies this statement.

Therefore it is recommended that we construct the less variable model so that the model is generalizable over the distinct sample sets. As described above, this suggestion can be achieved by using simpler and more parsimonious model. This recover the stability of the model in the expense of DA. Then, there remains an important question; How can we attain a balance between DA and DE? Or is it acceptable to allow a large value of DA to reduce DE so that the result model becomes replicable? The statement in the latter question may lead us to poor goodness of fit problem again. Consider the case

we use the model $\Sigma_k(\cdot) = I$, which is perfectly generalizable for any sample, but totally meaningless.

What we say 'generalizable' should not be understood in that manner. This word should embrace not only the smaller variation(DE) but also appropriate fit(DA). In this regard, overall discrepancy can e our best option in achieving our goal. Since OD is approximately equal to sum of DA and DE, this can act as a composite score of model accuracy and stability. In addition, The model with OD is supposed to be more interpretable since OD recover model parsimony compared with SD. Therefore, low value of OD can be a signal of a good model. Also, from now on we shall say 'generalizable' in the meaning of not only stability and less variability over distinct samples but also proper and degree of explanation on the interested phenomenon and practical interpretability.

Hence it is valuable to develop methods which lead us to the generalizable model in this respect. This can be done by modifying estimation methods which originally focus only on sample discrepancy. It will be highly worthwhile and significant if we implement a technique that eliminates unnecessary parameters in the estimation procedure so that the input model results in a more parsimonious one and produce less overall discrepancy. In fact, this is what we shall carry out in the present thesis, by means of 'Regularization', which will be illustrated in the next chapter.

Additionally, it also will be of help if some of fit indices that have been used in SEM can reflect overall discrepancy. If this is the case, we can use them as a guidance to the best generalizable and trustworthy model. ECVI can be an excellent option, since this index is defined in connection to the cross validation procedure. In fact,

$$
\begin{aligned}
ECVI = E(c) &\simeq E(F(\Sigma_0, \hat{\Sigma})) + \frac{p(p+1)}{2}/n \\
&= E(OD) + \frac{p(p+1)}{2}/n
\end{aligned}
\tag{1.4.1}
$$

Therefore, if we rank the model using in the order of ECVI value, it yields the same result with OD.

Since it is linearly related with ECVI, AIC also plays the same role. It is quite natural considering that this criteria is developed for the purpose of selecting generalizable model. BIC, which has the same objective with AIC, is worth consideration as well.

Many of absolute and comparative fit indices are not expected to perform the role since these indices focus on the fit in the current sample. Particularly most of comparative indices which are defined as a function of SD are not seem to work. Mc and RMSEA, which are part of absolute fit indices seem somewhat different since they are proposed as an estimator of DA. Note that SD is a biased estimator of DA(Steiger & Lind, 1980; McDonald, 1989). With some asymptotics, a less biased estimator of DA is suggested as

Even though Mc is proposed as an alternative to AIC in that the former is consistent over various samples but the latter is not(McDonald, 1989), what it's concerning is DA, not OD, since it is defined as a function of $\hat{F}_0$. RMSEA is also defined using $\hat{F}_0$, but there is a difference in tat RMSEA has a penalty term on model complexity. As a matter of fact, this intends to prevent the researcher from taking too complex model for the purpose of reducing DA. In other words, it is difficult to see that RMSEA concerns the generalizability issue.

Chapter 2

Regularization

2.1 Bias, Variance and MSE

In inferential statistics, one of the purposes researchers want to achieve is obtaining 'good' estimators for parameters which explain phenomena they are interested in. The term 'good' might imply various meanings, but mainly it indicates the 'accuracy' and 'stability' of estimators. And these two concepts are melted in the notion of 'MSE', the abbreviation for 'Mean Squared Error'.

In order to understand what MSE is, studies on concepts of loss and risk in statistics must take precedence. A loss is a function measuring the error, or the discrepancy between objects of estimation and their estimators(Wasserman, 2004; Hastie, Tibshirani & Freedman, 2008). There are lots of definitions of the loss, and some of them listed in Wasserman(2004) are as follows.

Various Definitions of the Loss

θ : Parameter $\quad \hat{\theta}$: Estimator of θ

i) Squared Error Loss $\quad L(\theta, \hat{\theta}) = (\theta - \hat{\theta})^2$

ii) Absolute Error Loss $\quad L(\theta, \hat{\theta}) = |\theta - \hat{\theta}|$

iii) L_p Loss $\quad L(\theta, \hat{\theta}) = |\theta - \hat{\theta}|^p$

iv) Zero-One Loss $\quad L(\theta, \hat{\theta}) = 0$ if $\theta = \hat{\theta}$ or 1 if $\theta \neq \hat{\theta}$

v) Kullback-Leibler Loss $\quad L(\theta, \hat{\theta}) = \int \log\left(\frac{f(x;\theta)}{f(x;\hat{\theta})}\right) f(x;\theta)dx$

The risk, or expected prediction/test/generalization error in accordance to the context or purpose of analysis, is defined as an expectation of a loss function(Wasserman, 2004; Hastie, Tibshirani & Freedman, 2008).

$$R(\theta, \hat{\theta}) = E\big[L(\theta, \hat{\theta})\big] \tag{2.1.1}$$

Note that risk can be defined in a various way according to what type of loss functions is used. MSE is one of the several definitions of the risk corresponding to the squared error loss(Wasserman, 2004; Hastie, Tibshirani & Freedman, 2008; Kim, 2012).

$$MSE(\theta, \hat{\theta}) = R(\theta, \hat{\theta}) = E\big[(\theta - \hat{\theta})^2\big] \tag{2.1.2}$$

Sometimes, when the purpose of research is prediction of Y using a fitted function $\hat{f}(X)$, an analogous concept of 'Mean Squared Prediction Error(MSPE)' is defined and studied(Kim, 2012).

The most well-known lemma on MSE is its decomposition into the sum of bias squared and variance of the estimator(Hastie, Tibshirani & Freedman, 2008; Kim, 2012).

$$
\begin{aligned}
MSE(\theta, \hat{\theta}) &= E\big[(\theta - \hat{\theta})^2\big] \\
&= E\big[((\theta - E(\hat{\theta})) + (E(\hat{\theta}) - \hat{\theta}))^2\big] \\
&= E\big[(\theta - E(\hat{\theta})))^2\big] + E\big[(\hat{\theta} - E(\hat{\theta}))^2\big] \\
&= \qquad Bias^2 \qquad + \quad Variance
\end{aligned}
\tag{2.1.3}
$$

In the above derivation, 'Bias' indicates a statistic that represent the difference between the object parameter and the expected value of its estimator. And 'Variance', as what it usually means, represent variability or standard error of the estimator. These two notions correspond with accuracy and stability of the estimator, respectively. Therefore, MSE can be understood as the measurement of how 'good' the estimator is, combining its accuracy and stability. From this point of view, MSE can be used as an index for generalizability of our estimator. The fact supports this statement that generalization error and expected generalization error are defined analogously to MSE in terms of their formulas.

Also, the basic logic inherent to this argument is exactly the same as what we discussed on several discrepancy concepts in Section 1.4. In fact, the concepts of MSE, Bias, and Variance are closely related to those of OD, DA, and DE in structural equation modeling. As illustrated, Σ_0 is the population covariance matrix, $\tilde{\Sigma}$ is the covariance matrix reproduced from the model estimated in population level, and $\hat{\Sigma}$ is the implied covariance matrix, which is the counterpart of $\tilde{\Sigma}$ in sample level. Conceptually, these matrices can be matched to θ, $E(\hat{\theta})$, and $\hat{\theta}$, respectively. Considering this matching, it is easy to understand the correspondence between risk-related indices(MSE, Bias, and Variance) and discrepancy indices in SEM(OD, DA, and DE).

The most noticeable difference between these groups of indices is that the

former is defined in a unit of each individual parameter, while the latter is in a unit of covariance matrix using discrepancy function, which is a matrix-valued function of those parameters. In this regard, MSE can be understood as a generalization index on the level of related parameters, while OD can be accepted as the same concept but on the level of covariance structure. Also, we shall use the term 'Matrix Discrepancy' to indicate matrix-based discrepancies such as OD, DA, and DE, and 'Parameter Discrepancy' to represent parameter-based discrepancies such as MSE, bias, and variance of estimators.

In sum, the two groups of indices share the same goal of measuring discrepancy between parameters and its estimators, but differ in the unit in their definition. However, the influence of this difference has not been studied and discussed much. This may be due to the reason that most researchers using SEM has been focusing on sample covariance, implied covariance, and model fit indices when evaluating their models. Also, since the estimation of SEM is carried out by means of the matrix discrepancy, characteristics of each estimator are hard to study. However, we are in opinion that appropriate attention should be paid to parameters in SEM. Estimation of parameters and related issues such as the case when model fits are excellent but some of the paths are insignificant should be studied more. And this can be a way to improve the quality of SEM estimation.

2.2 Shrinkage Estimation

In this section, a brief introduction to 'Shrinkage Estimation' will be given, as a stepping stone to understand the notion of 'Regularization' in linear regression analysis. Consider we want to estimate the unknown parameter $\mu = (\mu_1, \mu_2, \ldots, \mu_n)$ in the following problem.

$$\begin{aligned} Y_i &= \mu_i + \sigma\epsilon_i, \qquad i = 1, \ldots, n. \\ \epsilon_1, \ldots, \epsilon_n &\sim \ iid\ N(0,1) \end{aligned} \tag{2.2.1}$$

This kind of problem is called 'Many Normal Means Problem'(Wasserman, 2006). In this problem, the vector μ contains the same number of parameters as the sample size. That is, though it seems like a parametric problem, its complexity and features resemble the nonparametric problem. Note that, as the sample size is increasing, dimension of the parameter vector is becoming infinite-dimensional. The algebraic form of this problem is quite simple, but many of statistical problems can be expressed in this form. Among these problems are density estimation, nonparametric regression, etc(Wasserman, 2006).

The MLE estimator of this problem is $\hat{\mu}_i^{MLE} = Y_i,\ i = 1, \ldots, n$. This is an unbiased estimator of μ, and has the minimum variance among the estimators with unbiasedness. If we use this estimator, MSE, or the risk using the squared loss is computed as follows(Wasserman, 2006).

$$MSE(Y, \mu) = \sum_{i=1}^{n} E(Y_i - \mu_i)^2 = n\sigma^2 \tag{2.2.2}$$

This result is not that good considering the fact that this is proportional to the sample size n. Thus the risk is increasing as the sample size gets larger.

With regard to the many normal means problem in (2.2.1), some kind of biased estimators can be outstanding alternatives to MLE. The most famous

one is 'shrinkage estimators'. Consider the linear estimator in the following form.

$$\mu = bY, \quad 0 \leq b \leq 1. \tag{2.2.3}$$

This estimator shrinks the ML estimator Y toward zero. Among this kind of estimator, the following 'James-Stein Estimator' is the most widely known for the theorem stating that it nearly achieves the minimum risk called 'linear oracular risk'(Wasserman, 2006).

$$\hat{\mu}^{JS} = (1 - \frac{(n-2)\sigma^2}{\sum_{i=1}^{n} Y_i^2})Y. \tag{2.2.4}$$

The main point of this section is that there is a possibility that sometimes the established statistical method, including maximum likelihood estimation, cannot produce the optimal solution depending on properties of the current problem. In the example of many normal means problem, the MLE solution can be rated as quite good, but not the best. It is the most optimal only among the unbiased estimators. Breaking from the convention laying too much stress on unbiasedness, it is possible to obtain better solutions which are able to reduce MSE, the index of generalizability of the estimation.

In fact, this is what the modern statistics is doing. The next section will illustrate 'Regularization', which implements the same principle of the shrinkage estimation to linear regression models.

2.3 Regularization

In the multiple linear regression model, we attempts to predict the response variable, Y, using the regressor, X. Typically, the model is described as follows.

$$\begin{aligned} y &= X\beta + \epsilon \\ \epsilon|X &\sim iid\ N(0, \sigma^2) \end{aligned} \tag{2.3.1}$$

The OLS(Ordinary Least Squares) method aims to obtain the estimate of regression coefficient β such that minimizes the following loss function.

$$l(\beta) = \epsilon^T \epsilon = (y - X\beta)^T (y - X\beta) \tag{2.3.2}$$

Minimization of the loss function using differentiation yields the following equation, which is called 'Normal Equation' and its solution $\hat{\beta}_{OLS}$

$$(X^T X)\hat{\beta}_{OLS} = X^T y \tag{2.3.3}$$

$$\hat{\beta}_{OLS} = (X^T X)^{-1} X^T y \tag{2.3.4}$$

Since this sum of squared residuals loss function differs only in constant from the log-likelihood function, the MLE estimator is equal to that of OLS.

The model above implies four assumptions on the linear regression model; (1) Linearity, (2) Independence, (3) Homoskedasticity, and (4) Normality. Assumption (1) is quite natural as implied in the model itself. Assumption (4) is additional for the hypothesis testing of the significance of model and coefficients. Assumption (2) and (3) are called 'Gauss-Markov Assumption' along with the other assumption $E(\epsilon|X) = 0$, and these are the sufficient conditions for 'Gauss-Markov Theorem'. This theorem states that the OLS estimator of

the regression coefficient β is 'BLUE', short for 'Best Linear Unbiased Estimator'. This theorem implies that $\hat{\beta}_{OLS}$, the OLS estimator of β has the minimum variance among all the linear unbiased estimator of β.

Despite the great properties of BLUE, we still can ask the following question; is BLUE the best estimator among all the possible estimators? Traditional Statistics lean on this kind of belief. As described above, $\hat{\beta}_{OLS}$ can be understood as one of maximum likelihood estimators, which have a lot of great properties; namely, Unbiasedness, Efficiency, and Consistency. Furthermore, it is well known that MLE is one of the best estimators for achieving the 'Uniformly Minimum Variance Unbiased Estimator'(UMVUE) by applying several renowned theorems including 'Rao-Blackwell' and 'Lehmann-Scheffė' under some conditions(Kim, 2012). Of course, this group of estimators usually perform well enough to satisfy scientific researchers.

However, this properties is bounded only on the 'Unbiased Estimators'. Unbiasedness, which means the expectation of estimator is equal to its parameter, is an excellent property. Nonetheless it is not always the case that UMVUE becomes the best estimator minimizing MSE. Actually, state-of-the-art statistics focuses on minimizing MSE directly, rather than giving priority to unbiasedness. In fact, it is possible to allow a little bias for the sake of reducing much more variance of the estimator.

As described in the previous section, in many cases which can be transformed to namely 'many normal means problem', the shrinkage estimators act as the cutting-edge of this spirit. Also some techniques that perform the same role in the linear regression model have been developed. These techniques, which are called 'Regularization', share the property to shrink the absolute values of the estimates of regression coefficients toward(or to, sometimes) zero. This property have the same effect as the shrinkage estimators in many normal

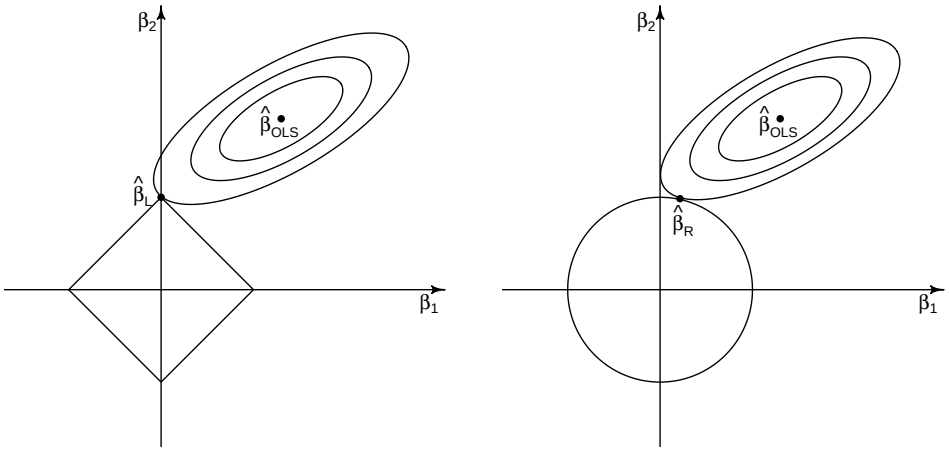

Figure 2.1: Regularization

means problem; reduce MSE by diminishing variance largely, at the expense of the unbiasedness.

Regularization can be understood as 'imposing constraints on the parameter space'. In figure 2.1, the contour plots indicate the OLS loss function in equation (2.3.2). The value of the function depends on two regression coefficient parameters, β_1 and β_2. This value is minimized at the point in the middle of the contours, where the OLS estimates are obtained. Thus, the estimates of β_1 is the value of the foot of perpendicular from the minimum point on the horizontal axis. The estimates of β_2 can also be obtained analogously.

Now consider the case that constraints are imposed on the parameter space like the diamond form for the left figure, and the round form for the right one. By 'imposing constraints', it is meant that the estimates of parameters are forced to be obtained within the given space. Therefore, the new estimates are drawn at the point where the contours and the restricted parameter space are tangent to each other. Regardless of types of the constraints imposed on the parameter space, it should be noted that values of the new estimates are

all less than those of the original OLS estimates – they are shrunk to/toward zero. Obviously, this is the reason for categorizing the regularization as a kind of shrinkage estimation.

The followings are some details on each of regularization methods, part of which are related to figure 2.1.

2.3.1 Ridge (Hoerl & Kennard, 1970a, b)

Ridge is one of regularization methods that imposes a rounded constraint on the parameter space as the left in the figure 2.1. Sum of squared beta coefficients penalty in equation (2.3.5) implies this kind of constraint.

$$l(\beta) = (y - X\beta)^T(y - X\beta) + \kappa \sum_{j=1}^{p} \beta_j^2 \tag{2.3.5}$$

κ is a tuning parameter, which controls degree of restriction on the parameter space.

This method was proposed as a remedy to multicollinearity among the regressor variables. When a new regressors are added to the original regression model, standard errors of β_{OLS} increases in proportion to the degree of linear dependencies among these regressors. That is, when the regressor variables show high linear correlations, variances of the estimators are severely inflated. This is related to the fact that X^TX becomes ill-conditioned, which means the matrix is close to singularity, in the case of multicollinearity. This makes it difficult to obtain inverse of X^TX, also the OLS estimates $\hat{\beta}_{OLS} = (X^TX)^{-1}X^Ty$.

As a countermeasure to the multicollinearity case, consider the following equation and its solution $\hat{\beta}_{Ridge}$(Hoerl & Kennard, 1970a, b; Montgomery, Peck & Vining, 2012). Note that the equation can be obtained by differentiating (2.3.5).

$$(X^T X + \kappa I_p)\hat{\beta}_{Ridge} = X^T y \tag{2.3.6}$$

$$\hat{\beta}_{Ridge} = (X^T X + \kappa I_p)^{-1} X^T y \tag{2.3.7}$$

Constant k is selected by researchers using what is called 'Ridge trace'. Ridge solution equals to OLS estimator when $k = 0$. Re-expressing $\hat{\beta}_{Ridge}$ shows its relationship to OLS solution appears more obviously(Hoerl Kennard, 1970a).

$$\begin{aligned}
\hat{\beta}_{Ridge} &= (X^T X + \kappa I_p)^{-1} X^T y \\
&= (X^T X + \kappa I_p)^{-1} (X^T X)(X^T X)^{-1} X^T y \\
&= (X^T X + \kappa I_p)^{-1} (X^T X)\hat{\beta}_{OLS} \\
&= (I_p + \kappa X^T X)^{-1} \hat{\beta}_{OLS}
\end{aligned} \tag{2.3.8}$$

From this expression, several properties of Ridge estimators are revealed(Hoerl & Kennard, 1970a; Montgomery, Peck & Vining, 2012).

i) Ridge estimator is a biased estimator. its bias is continuous, and monotonically increasing as k increases.
ii) Variance of Ridge estimator is continuous, and monotonically decreasing as k increases.
iii) There exists a nonzero k such that makes MSE($\beta, \hat{\beta}_{Ridge}$) less than MSE($\beta, \hat{\beta}_{OLS}$) unless $\beta^T \beta$ is infinite, which is clear in practice.

Hence, Ridge method enables its users to attain an estimator more stable than that of OLS method.

Another strength of Ridge is its computational easiness. Since Ridge implement a squared, or L_2 penalty as can be seen in (2.3.5), Ridge's loss function

is differentiable which makes a closed form solution available. This solution is shown in (2.3.7). Futhermore, this solution can be computed directly as an OLS solution. Consider the following matrix and vector obtained by augmenting the original X and y.

$$X_A = \begin{pmatrix} X \\ \sqrt{\kappa} I_p \end{pmatrix}, \qquad y_A = \begin{pmatrix} y \\ 0_p \end{pmatrix} \tag{2.3.9}$$

Inputting these into the OLS formula (2.3.4), $\hat{\beta}_{Ridge}$ is computed directly.

$$(X_A^T X_A)^{-1} X_A^T y_A = (X^T X + \kappa I_p)^{-1} X^T y \tag{2.3.10}$$

2.3.2 Lasso (Tibshirani, 1996)

As another regularization method, Lasso was proposed by Tibshirani(1996). Lasso is the abbreviation for 'Least Absolute Shrinkage and Selection Operator'. As this name implies, this method implements an absolute-valued penalty, or L_1 penalty to the loss function.

$$l(\beta) = (y - X\beta)^T (y - X\beta) + \kappa \sum_{j=1}^{p} |\beta_j| \tag{2.3.11}$$

L_1 penalty forms a diamond-shaped constraint on the parameter space as shown in the left figure in 2.1. This geometric feature produce Lasso's Most attractive property; when OLS loss function meet the diamond constraint region at its corner, some coefficients in the current model shrink completely to zero. Compare this property with that of Ridge, which also shrinks the coefficients toward zero, but not completely to zero. When the complete shrinkages occur,

variables corresponding to those zero coefficients may be regarded as being deleted from the original model. Therefore Lasso conduct both the shrinkage estimation and variable selection as the title of Tibshirani(1996) suggests.

Due to the absolute constraint, Lasso doe not have a closed form solution. Instead, since the minimization problem of equation (2.3.11) can be viewed as a linearly constrained quadratic programming problem, various solution methods to this kind of problem can be exploited. Among these are conjugate gradient method, simplex, and so on.

Especially for Lasso, Least Angle Regression or LARS can be used as an extremely fast and efficient estimating algorithm(Efron et al., 2004). LARS is a modified version of 'Forward Stagewise Linear Regression' that improves the computing speed remarkably. Even though their fundamental logic is quite distinct, a slight modification of LARS yields almost exactly the same estimates as those of Lasso. A brief introduction to LARS will be illustrated in the following section.

In addition, MM-algorithm, one of general optimization methods, is also able to produce Lasso solution. This is a generalized version of EM-algorithm, which can be used in various minimization/maximization problems. When applied to Lasso's loss function, it yields an iterative solution formula converging to Lasso estimates. This will also be introduced with LARS.

In spite the fact that generally no closed form solution is available for Lasso, it is possible to obtain it when the design matrix X is orthogonal. When this is the case, simply differentiating the function (2.3.11) and dividing the cases regarding the sign of OLS coefficients, the following solution can be derived.

$$\hat{\beta}_{Lasso} = sign(\hat{\beta}_j)(\hat{\beta}_j - \kappa)_+, \quad \hat{\beta} = \hat{\beta}_{OLS} \tag{2.3.12}$$

This is the same as the soft-thresholding estimator, whose risk is near to the

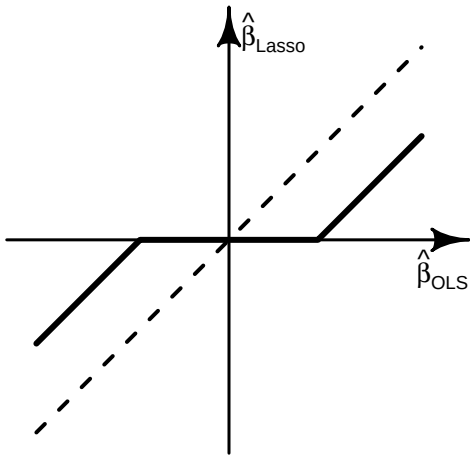

Figure 2.2: Soft-thresholding

oracular case in wavelet methods(Donoho and Johnstone, 1994; Wasserman, 2006; also, Tibshirani, 1996; Hastie, Tibshirani Friedman, 2008). Figure 2.2 represents this estimator and may be of help to understand Lasso.

At last, note that Lasso and Ridge can be grouped into the 'Bridge', which was proposed by Frank and Friedman(1993). They illustrated some properties of the following type of regularization, for various values of γ. $\gamma = 1$ and $\gamma = 2$ yield Lasso and Ridge, respectively.

$$l(\beta) = (y - X\beta)^T(y - X\beta) + \kappa \sum_{j=1}^{p} |\beta_j|^\gamma \tag{2.3.13}$$

2.3.3 Elastic Net (Zou & Hastie, 2005)

Elastic Net is a hybrid method of regularization combining Ridge and Lasso. This description evidently appears in the following loss function.

$$l(\beta) = (y - X\beta)^T(y - X\beta) + \kappa_1 \sum_{j=1}^{p} |\beta_j| + \kappa_2 \sum_{j=1}^{p} \beta_j^2 \tag{2.3.14}$$

The penalty term in (2.3.14) is equal to the weighted sum of Ridge and Lasso penalties. This enables Elastic Net to have strengths of both Ridge and Lasso. That is, as L_2 penalty corrects the multicollinearity and enjoys its grouping effect, L_1 penalty performs the variable deletion at the same time(Zou & Hastie, 2005; Kyung, 2014).

Zou and Hastie(2005) also showed that minimization of (2.3.14) is equivalent to the penalized least squared problem which minimizes residual sum of squares, subject to $\left[(1-\alpha)\sum_{j=1}^{p}|\beta_j| + \alpha\sum_{j=1}^{p}\beta_j^2\right] \leq t$ for some t, where $\alpha = \kappa_2/(\kappa_1 + \kappa_2)$. Note that letting $\alpha = 0$ or $\alpha = 1$ yields Lasso or Ridge penalties, respectively.

Additionally They proved that, in a similar way to the augmenting method in Ridge case, (2.3.14) can be transformed to Lasso-type loss function. This enables Elastic Net problem to be optimized by several computing methods developed for Lasso, including LARS.

It should be noted that, In fact, a solution to (2.3.14) suffers from over-shrinkage problem. That is, shrinkage effect is doubled since the Elastic Net loss function implements two distinct penalty terms. This is the reason why Zou and Hastie called this method as 'Naive' Elastic Net. However, they also proposed a rescaling method to remove this double-shrinkage effect preserving each of nice properties in Ridge and Lasso. And rescaled solution is named Elastic Net estimate.

Even though geometrical forms of constraints are indistinguishably similar, which seem like a rounded diamond, Elastic Net performs quite differently from L_γ regularization, where $1 \leq \gamma \leq 2$ (Hastie, Tibshirani & Fried-

man, 2008). Unlike L_γ regularization, the constraint region of Elastic Net has sharp corners which enables to shrink some coefficients completely to zero, like Lasso regularization. L_γ constraint has a rounded corner so that it does not has this property. Interested readers may find more details with figure of constraints of Elastic Net and L_q regularization in Section 3.4.4 in Hastie, Tibshirani and Friedman(2008).

2.4 The Connection between Regularization and Bayesian Analysis

One of the interesting facts on shrinkage estimation and regularization is that the Bayesian Approach can be understood as this kind of estimation. Before describing this relationship, we review a simple example of Bayesian Linear Regression Analysis.

2.4.1 Bayesian Linear Regression Analysis

Consider the following semiconjugate prior distribution for regression coefficient β and the variance σ^2(Hoff, 2009).

$$\begin{aligned} \beta &\sim N_p(\beta_0, \Sigma_0) \\ \sigma^2 &\sim Inverse\text{-}Gamma(\nu_0/2, \nu_0\sigma_0^2/2) \end{aligned} \tag{2.4.1}$$

And with the model (2.3.1), we can obtain the following likelihood function.

$$p(y|X, \beta, \sigma^2) = (\frac{1}{\sqrt{2\pi\sigma^2}})^n \exp(-\frac{1}{2\sigma^2}(y - X\beta)^T(y - X\beta)) \tag{2.4.2}$$

It is well known that posterior distributions in Bayesian analysis can be derived by the formula $p(\theta|y) \propto p(\theta) \times p(y|\theta)$. Combining the priors with the normal likelihood function from our model, the next posterior distributions can be derived.

$$\begin{aligned} &\beta|y, X, \sigma^2 \sim N_p(\beta_n,\ \Sigma_n) \\ &\quad \text{where } \Sigma_n = (\Sigma_0^{-1} + X^TX/\sigma^2)^{-1}, \\ &\qquad \beta_n = (\Sigma_0^{-1} + X^TX/\sigma^2)^{-1}(\Sigma_0^{-1}\beta_0 + X^Ty/\sigma^2) \\ &\sigma^2|y, X, \beta \sim Inverse\text{-}Gamma((\nu_0 + n)/2,\ (\nu_0\sigma_0^2 + SSR(\beta))/2) \\ &\quad \text{where } SSR(\beta) = (y - X\beta)^T(y - X\beta) \end{aligned} \tag{2.4.3}$$

Other examples on the basic bayesian linear regression can be found in Gill(2007), Gelman et al.(2014), etc.

It should be noted that the likelihood function is combined with given priors in computing posterior distributions. This step can be understood in a different way; that is, the estimation is conducted using the likelihood function, like MLE, but under the constraint presented in the priors. This claim implies that Bayesian linear regression can be interpreted as a kind of regularization. Or regularization can be understood as a special case of Bayesian analysis. The following clarifies this argument. Here, the likelihood is from the normal density as in (2.3.1).

$$\begin{aligned} p(\beta|y) &\propto p(\beta) \times p(y|\beta) \\ &\propto \exp(-\frac{1}{2}(y - X\beta)^T(y - X\beta)) \times p(\beta) \\ &(\Rightarrow \text{ Taking logarithm on the RHS }) \\ &\Rightarrow \ (y - X\beta)^T(y - X\beta) + c\log(p(\beta)) \\ & \qquad\qquad\qquad\qquad c : \text{constant.} \end{aligned} \tag{2.4.4}$$

Note that the equation in the last line is in the same form with those of regularization methods (2.3.5), (2.3.11), and (2.3.14).

In addition, some studies on Ridge regression (2.3.5) can also be viewed as an example supporting the argument. Consider the posterior mean of β in (2.4.3), with hyperparameters in the prior as $\beta_0 = 0, \Sigma_0 = \frac{\sigma^2}{\kappa} I_p$. Then,

$$\begin{aligned} \hat{\beta} = \beta_n &= ((\frac{\sigma^2}{\kappa} I_p)^{-1} + X^T X/\sigma^2)^{-1}(\frac{\sigma^2}{\kappa} I_p \times 0 + X^T y/\sigma^2) \\ &= ((\frac{\kappa}{\sigma^2} I_p) + X^T X/\sigma^2)^{-1}(X^T y/\sigma^2) \\ &= (X^T X + \kappa I_p)^{-1} X^T y \end{aligned} \tag{2.4.5}$$

The result equals to $\hat{\beta}_{Ridge}$ in (2.3.7). This shows the close connection between Ridge estimator and bayesian linear regression(Montgomery, Peck & Vining, 2012). Similarly, Ridge type of regularization can be conducted as Bayesian analysis, using the following priors.

$$\begin{aligned} \beta &\sim N_p(\beta_0, \frac{\sigma^2}{\kappa} I_p) \\ \sigma^2 &\sim Inverse\text{-}Gamma(\nu_0/2, \nu_0\sigma_0^2/2) \end{aligned} \tag{2.4.6}$$

Again, this shores up the statement claiming the connection between regularization and Bayesian analysis.

2.4.2 BLasso: Bayesian Lasso

Also, other regularization methods we described above can be carried out with a Bayesian Approach. For example, a Bayesian version of Lasso has widely been studied during the past decades(Park & Casella, 2008; Hans, 2009). These studies were stimulated by Tibshirani's comment left on his previous paper proposing Lasso. What he said is that Lasso estimates can be derived as 'the Bayes posterior mode under independent double-exponential priors' which equals to the following(Tibshirani, 1996).

$$f(\beta_j) = \frac{\kappa}{2}\exp(-\kappa|\beta_j|) \tag{2.4.7}$$

Park and Casella(2008)'s BLasso can be considered as one of the most successful studies on Bayesian approach to Lasso regularization. Their strategy is to use the conditional Laplace prior for regression coefficient β and the non-informative scale-invariant marginal prior for variance σ^2 (Park & Casella,

2008).

$$\begin{aligned} \pi(\beta|\sigma^2) &= \prod_{j=1}^{p} \frac{\kappa}{2\sqrt{\sigma^2}} \exp(-\frac{\kappa|\beta_j|}{\sqrt{\sigma^2}}) \\ \pi(\sigma^2) &= 1/\sigma^2 \end{aligned} \tag{2.4.8}$$

They revealed that these priors can be transformed to another expression, which is much easier to be applied to the actual estimation procedure. Using the representation of the Laplace distribution as a scale mixture of normal distribution with an exponential mixing density, Park and Casella(2008) suggested the 'Hierarchical Representation' of the prior (2.4.8), which shows the conjugacy between prior and posterior distribution(Andrews & Mallows, 1974; Park & Casella, 2008).

$$\begin{aligned} y|X,\beta,\sigma^2 &\sim N(X\beta,\sigma^2 I_n) \\ \beta|\sigma^2,\tau_1^2,\ldots,\tau_p^2 &\sim N_p(0,\sigma^2 H), \;\; H = diag(\tau_j^2)_{j=1,\ldots,p} \\ \sigma^2 &\sim \pi(\sigma^2)d\sigma^2 \\ \tau_j^2 &\sim \exp(\frac{\kappa^2}{2}),\; j=1,\ldots,p \\ \sigma^2,&\tau_1^2,\ldots,\tau_p^2 > 0 \end{aligned} \tag{2.4.9}$$

Here, $\pi(\sigma^2)$ can be any distribution for σ^2 which can play a role as the conjugate prior. Bayesian analysis using these priors yield nearly the same effects to Lasso regularization.

However, several limitations should be pointed out with respect to the Bayesian approach to Lasso. At first, as Park and Casella said, the BLasso's estimates are not shrunk exactly to zero(Park & Casella, 2008). They used Bayesian posterior medians as their estimates. And in the figure comparing BLasso with Lasso and Ridge(figure 2.3), though BLasso's trace plot may seem similar to that of Lasso, they revealed that the result didn't show any sparsity. Perhaps

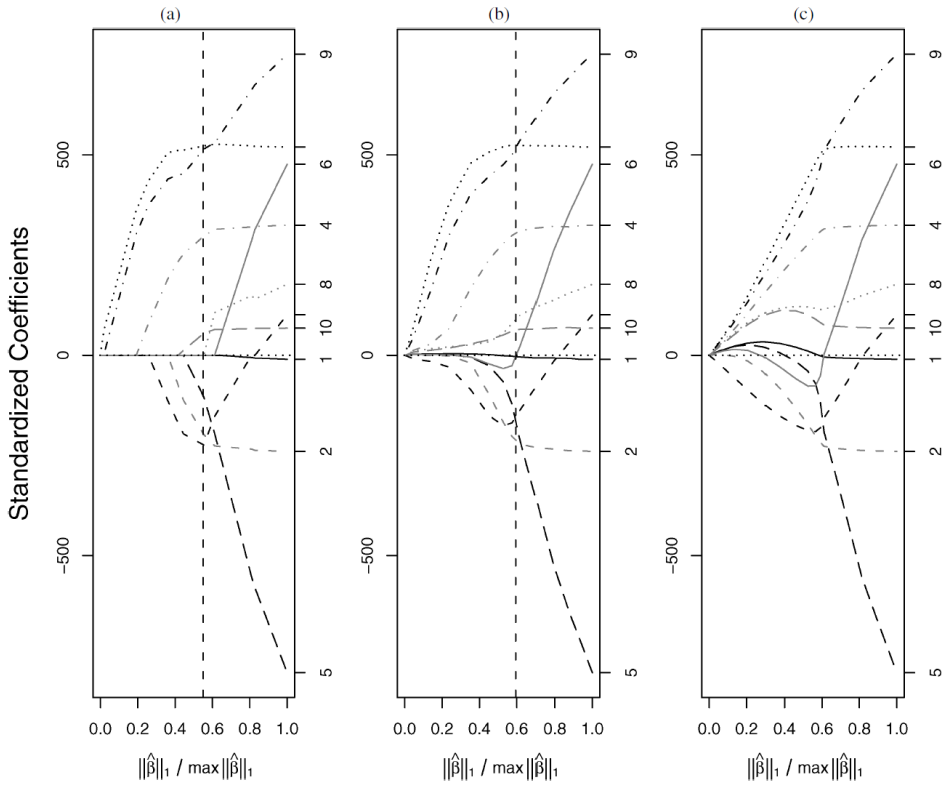

Figure 2.3: Trace plots of a) Lasso, b) BLasso, and c) Ridge in Park & Casella(2008).

this is due to the procedure that the Bayesian estimates are obtained from. In Bayesian analysis, we derive or approximate posterior distributions and use its mean or median as our estimates. To obtain its value, several sampling methods are exploited to get samples from the posterior distributions. And the sample mean or the sample median are computed from this sample. However, it is hard to expect these values to be zero owing to several factors including sampling error. Additionally, Even when the exact form of posterior distribution is available, there is no guarantee that the estimates become exactly zero.

This may not be a big problem. Some alternative strategies can be exploited to decide whether the interested coefficients are shrunk completely to zero or

not. As an example, we may determine a bound in advance, and regard the coefficients as being zero if their values drop within the bound. Some times this kind of rule of thumb can be of help, even though they have no theoretical backgrounds.

The second point is that, it seems that BLasso acts as a compromise of the other two – Ridge and Lasso. Park and Casella mentioned that 'the paths(of BLasso) are smooth, like ridge regression, but are more similar in shape to the Lasso paths, particularly when the L_1 norm(= $||\hat{\beta}||_1/\max||\hat{\beta}||_1$) is relatively small'(Park & Casella, 2008). However, other interpretations are also possible on the same figure. In Lasso's trace plot, coefficient are shrunk completely to zero in sequence. BLasso's trace plot shows that its estimates seem to converge to zero together, only when $||\hat{\beta}||_1/\max||\hat{\beta}||_1$ is close enough to zero. This tendency also appears in Ridge's trace plot. This is because BLasso and Ridge are not capable of shrinking coefficients to zero perfectly. The trend shows us the fact that in Ridge and BLasso, coefficients become trivial only when the tuning parameter is very large, but still not zero.

In addition, the BLasso estimates determined based on marginal maximum likelihood method, the technique suggested as a tool for choosing the tuning parameter in Park and Casella(2008), does not include any zero coefficient as we pointed out previously. Only one coefficient seems to be close to zero in the figure. This offers another big difference between trace plots of Lasso and BLasso.

These two points are closely related to the fact that BLasso does not show the complete shrinkage effect. Note that this effect is the most notable feature of Lasso. Thus, without this property, we can hardly consider the BLasso's result as essentially similar to that of Lasso.

The last point deserves much greater attention. Consider the case that BLasso

has some exact zero estimates and this appears in sequence according to their importance in explaining or predicting the value of response. This is parallel to a normal Lasso result. However, in case we conduct a Bayesian analysis, this estimates does NOT mean that the corresponding variables are deleted. In fact, we should interpret this case as these estimates have distributions whose means or medians are equal to zero. Existence of these distributions never means that researchers allowed to remove them. Therefore even when BLasso yields the same result as that of Lasso, we are not able to enjoy the variable deletion and model selection property.

In conclusion, despite its well-known strengths over almost the whole ranges of statistical analysis, Bayesian approach does not seem to be a very good match for Lasso-type regularization.

2.5 Regularization and Structural Equation Modeling

Among the above techniques, Lasso and Elastic Net have been receiving huge attention from the time when they are proposed. This is due to the fact that this techniques are able to select the model naturally in the process of parameter estimation. Unlike Ridge, which only shrink the estimates 'toward' zero, other two methods can shrink some of them completely 'to' zero. Regressor variables corresponding to the coefficients shrunk to zero can be regarded as being 'removed' from the original model. Therefore, researchers who make use of these tools can avoid the trouble of choosing the best model amongst all the possible candidates.

This complete shrinkage effect has huge significance in various aspects. At first, techniques with this property is greatly effective in the regard that they can estimate and select the result model at the same time. Thus, scientific researchers don't have to rely on several existing model selection techniques based on 'data-drivening' which make them vulnerable to characteristics intrinsic only to the current sample data. This point was well-described before.

Another strength of this kind of technique is it's capability to produce a much simpler model as a result. This feature, which is related to 'Sparsity' or 'Parsimony' of models is especially great in science. In a scientific research, the result of analysis is important per se, of course. However, what is far more important for researchers are its interpretation and implication. Sometimes results without, or with less interpretability are devaluated since they have a little practical significance, particularly in the fields where the main purpose of science is an explanation of the interested phenomenon. Therefore, sparse and parsimonious models can be more valuable in a regard that they are far more useful for us to make an scientific interpretation, comparing to more complex

one.

Sparsity and Parsimony is important in another aspect. Consider the next example, which is originally introduced in Bishop(2006). we have 10 data points sampled from a population function which has cubic trend, namely $f = \sin(2\pi x)$ as in the original example. And we fit the following polynomial regression model to fit the data.

$$y = \beta_0 + \sum_{j=1}^{D} \beta_j x^j + \epsilon \tag{2.5.1}$$

Since we have no idea on the population trend, determination of dimensionality D is an important issue. $D = 3$ may make the result model most closest to the tendency in population(the leftmost in figure 2.4), but other values can also be good alternatives. In fact, in case of $D > 3$ the model yields less residual sum of squares and R^2 than $D = 3$. And in addition, residual sum of squares becomes zero and R^2 becomes one when $D = 9$. As a matter of fact, $D = 9$ yields the fitted line that passes through all the 10 points in the sample(the middle in figure 2.4). This means in the current sample data, $D = 9$ can be regarded as a perfect solution. However, since this 9-th degree curve has

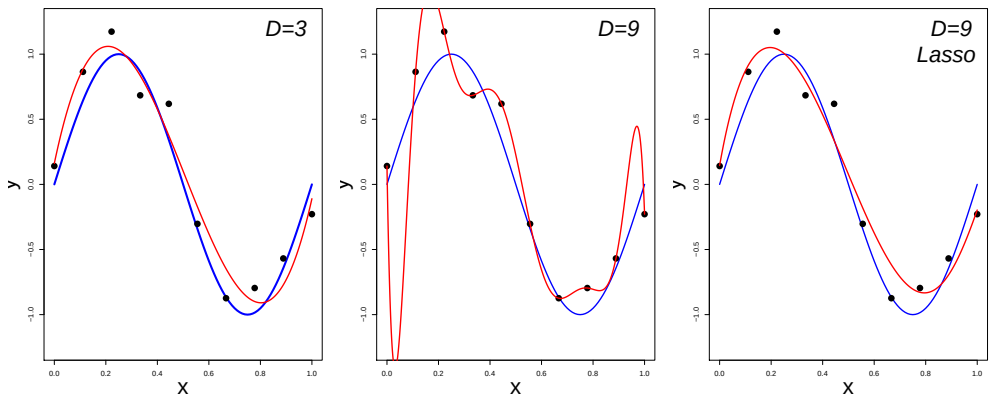

Figure 2.4: Overfitting Problem and Lasso

$9 - 1 = 8$ inflection points, it oscillates too severly, unlike the near 3-th degree trend shown in the data points. In other words, even though the fitted curve has perfect goodness of fit, it can not be accepted as an appropriate model estimation result. This problem is called 'Over-fitting' which appears when the model is too complex for the sample size of given data.

The rightmost panel in figure 2.4 represents the Lasso result. In this toy-example, Lasso yields 5 zero coefficients when the model with $D = 9$ is fitted. Nonzero coefficients come out for $j = 1, 2, 3,$ and 8. And this result retrieve nearly the same trend with that of the leftmost case. Futhermore, these two curves, $D = 3$ and $D = 9$ with Lasso, trace almost exactly the population trend. Hence, it seems proper that we conclude Lasso can be an effective tool for alleviating the overfitting problem.

Furthermore, the overfitting issue is related to generalizability of statistical models. Sparser and more parsimonious models can enjoy less variance(standard error) of each coefficients, and less variability in the model fitting result itself. And this yields more stable results when the same model is fitted over many distinct sample data sets. In the above polynomial regression example, a simpler model shows less prediction error. Also when it comes to the case that analysis purpose is on explanation of scientific phenomenon, a simpler model tends to perform better than a complex one, even though goodness of fit computed with the original data is better in the latter than in the former. This implies the most important point in the present thesis; complete shrinkage effect may lead us to more generalizable models that predict or explain the future well.

From this point of view, we can find possibility of improving the structural equation modeling by means of the techniques with complete shrinkage effect. By 'improving' we intend to mean that these techniques may enhance the

generalizability of SEM. Note that SEM is a system of many equations with linear regression forms. Thus, it may be able to extend this simultaneous model estimation and variable selection effect in Lasso or Elastic Net to SEM.

It should be noted that fitting with Lasso or Elastic Net may depress the conventional goodness of fit indices. However, most of these indices are related to sample discrepancy, the counterpart of residual sum of squares or R^2 in linear regression analysis. As we saw with the example, minimizing this kind of criterion does not guarantee the result to be repeated stably over distinct samples – it just produces a result best only in the current sample. Even though some of traditional fit indices in SEM have penalty on model complexity, most of them are not designed to trace the overall discrepancy(OD), a discrepancy representing generalizability in SEM. Also, these indices are not fully tested with respect to generalizability(of course, some good exemplary studies exist such as Preacher et al.(2013)).

The discussion so far is related to the reproducibility issues in psychological researches, particularly for which use SEM to support researchers' theories and hypotheses. Psychologists use SEM when they attempt to make models explaining relationships among the variables with some of them are unobserved and latent. But researchers are never able to know thoroughly the true model or the true data-generating process in the population. Thus, even though they set up the model carefully based on well-established theoretical background, the result might be poor or not represent the true covariance structure well – actually, it is uncommon to observe the good result in scientific researches.

Therefore, to improve model fit, sometimes we fit a number of models to the same data or add some paths or variables to the original model, in spite that these models, paths and variables' theoretical backgrounds are a little frag-

ile. Similar things occur when we attempt to develop the existing, validated model.

The result may seem good in a regard of goodness of fit, but naturally are vulnerable to the problem of overfitting and data-drivening. This kind of effort to obtain the 'good' result in the current sample may produce a contribution to psychology. However, if they are excessive, these efforts can be obstructions to other efforts to keep the status of psychology as a field of science such as trying to enhance the reproducibility of psychological researches.

As a remedy to extenuate these problems related to reproducibility, we attempt to regularize the structural equation models via the Lasso, one of the regularization methods able to make some nuisance path coefficients shrink completely to zero. If L_1 penalty works well for SEM, as it does in linear regression models, this effect enable us to obtain a sparser and more parsimonious model. Furthermore, the result model has relatively low model complexity and diminished potential risk to fall in overfitting problem so that it is able to maintain its valid explanation when it fits to other independent test samples, comparing to the original input model containing several nuisance parameters and having higher complexity. That is, Lasso can produce more reproducible and replicable results for SEM.

There is a noteworthy point on SEM and generalizability issue. In the example of linear regression and overfitting problem, we mentioned regularization and generalizability of the model with respect to its prediction error. That is, a generalizable model indicates the one good at predict future observations. However, SEM does not aim at prediction. Since most of regressors in the model is latent, basically the model cannot be used in expecting future data. Instead, SEM put its purpose on explanation of scientific, especially psychological phenomenon. If the model fits well in a sample, but poorly in other sample

sets, the theory implied in this model cannot be supported and accepted scientifically. On the contrary, if the model fits well over many distinct sample sets from the same population, this model is able to give us an insight into the real world tendency in the population. In this light, for SEM and other models/methods/experiments whose purpose is on explanation, generalizability should be understood in the context of reproducibility and replicability.

Someone may claim that Elastic Net can be another option. We also agree on that opinion, but its penalty is rather more difficult to deal with, than that of Lasso. Since SEM is far more complex model than ordinary linear regression model, this unmanageable penalty may yield some complex hassles with respect to its tuning/optimization. In addition, Elastic Net's merits over the Lasso – according to Zou and Hastie(2005), grouping effect, stability in $n << p$ case, and so on, are not that attractive in SEM; each equation in SEM is uncommon to have regressors that should be treated as a group and cases that the number of regressors are more than the sample size. Since our main purpose is on the estimation of SEM with parsimony and sparsity, Lasso can be regarded as a great starting point in regularizing the SEM.

As a matter of fact, there were already some efforts to regularize SEM though their purposes was different from that of the present thesis. We can use the Ridge estimation in LISREL, one of the leading programs in analyzing SEM. This just add a minor constant diagonal matrix to sample covariance matrix S as a remedy to non-positive definite S case(Jőreskog & Sőrbom, 1996; Yuan, Wu & Bentler, 2011). However, as illustrated above, Ridge shrinks some coefficients only toward zero, not completely to zero. Therefore, this method is not appropriate to achieve our purpose of obtaining a sparser and more parsimonious result and improving generalizability of the model.

Furthermore, there exist several studies on regularizing SEM indirectly by

applying Bayesian approaches(Guo et al., 2012; Wang, Z., 2014). Among them, application of Bayesian Lasso to SEM will be reviewed in Chapter 3. However, as we pointed out in Section 2.4, Lasso regularization with Bayesian approach have several problems – estimates of nuisance coefficients do not converge perfectly to zero, its performance seems like the compromise of Ridge and Lasso, and as the most severe one, even though an estimate shrinks completely to zero, it's just a distribution with zero central tendency which cannot be regarded as being deleted from the model. Since the last point denies the most critical feature of Lasso-type regularization, it needs to be discussed continuously henceforth. These limitations implies the significance of implementing Lasso to SEM in a more direct way.

Before we go to the next chapter, we shall illustrate some optimization methods that can be exploited to Lasso regularization, as a foundation stone to extend it to the structural equation modeling.

2.6 Some optimization methods for Lasso

2.6.1 LARS Algorithm

LARS(Efron et al., 2004; Hastie, Tibshirani & Friedman, 2008) is one of the most famous algorithm to fit linear regression models in statistics. A distinction between original OLS estimation and LARS is drawn from its capacity of estimating and selecting a model simultaneously. In fact, this algorithm is introduced as a kind of hybrid method of Lasso and 'Forward Stepwise Regression', which is one of classic model selection techniques.

Before start, assume that the response is centered and regressors are standardized as in Efron and colleagues' previous paper.

$$\sum_{i=1}^{n} y_i = 0, \quad \sum_{i=1}^{n} x_{ij} = 0, \quad \sum_{i=1}^{n} x_{ij}^2 = 1 \tag{2.6.1}$$

Let $\hat{\beta}$ be a p-length vector of estimates for linear regression coefficients. Then $\hat{y} = X\hat{\beta}$ yields our predicted values for the response vector y. While OLS derives its estimates by minimizing a residual sum of square $S(\hat{\beta}) = \sum_{i=1}^{n}(y-\hat{y})^2$, LARS obtain its solution by means of its own process, which Efron et al. called 'Equiangular Strategy'. LARS estimate vector $\hat{y}$ begins with $\hat{y} = 0$ which indicates that no regressor variables included in the estimate. If y is not centered before the procedure, $\hat{y} = \bar{y}$ is chosen as the starting point.

At the first step, the algorithm choose a variable that has the largest absolute value of correlation with a residual vector. This selection is carried out by the magnitude of the 'Current Correlation'

$$\hat{c} = c(\hat{y}) = X^T(y - \hat{y}) \tag{2.6.2}$$

The variable selected in the current step, namely x_{j1} is included in the 'Active Set' $\mathcal{A}$. Now let $\hat{C}$ be the maximum value of current correlations. Then this

procedure can be presented as follows.

$$\begin{aligned} &\hat{C} = \max_j \{|\hat{c}_j|\}, \quad \mathcal{A} = \{j : |\hat{c}_j| = \hat{C}\} \\ &\hat{c}_j : \text{current correlation of } j\text{-th variable.} \end{aligned} \tag{2.6.3}$$

We update our estimate in the direction of x_{j1}. Degree of updating is determined so that some other variable shows the same magnitude of current correlation with x_{j1}. Denote this variable as x_{j2}. In the subsequent step, this x_{j2} is included in the active set and updating occurs in the direction which has the equal angle with x_{j1} and x_{j2}.

The updating continues in the same manner. That is, the new estimate is determined by moving the original one in the equiangular direction of the variables in the active set. This procedure can be described as follows.

$$\tilde{y}(\gamma) = \hat{y}_{\mathcal{A}} + \gamma u_{\mathcal{A}} \tag{2.6.4}$$

$$\hat{y}_{\mathcal{A}_+} = \hat{y}_{\mathcal{A}} + \hat{\gamma} u_{\mathcal{A}} \tag{2.6.5}$$

where $\hat{y}_{\mathcal{A}}$ is a current estimates and $u_{\mathcal{A}}$ is an equiangular vector which guides the updating to the direction dividing into equal angles among the variables in the active set. And γ indicates the size of the updating step whose estimate $\hat{\gamma}$ is computed at each step. Here, computation of $u_{\mathcal{A}}$ and $\hat{\gamma}$ becomes the issue.

For $u_{\mathcal{A}}$, let $A_{\mathcal{A}}$ be the equiangle value we need to obtain $1_{\mathcal{A}}$ be a vector consisting of 1's whose length is equal to the number of elements in $\mathcal{A}$. And also let $X_{\mathcal{A}} = (\ldots, s_j x_j, \ldots)_{j \in \mathcal{A}}$ where $s_j = \text{sign}(\hat{c}_j)$ for $j \in \mathcal{A}$. We define the equiangular vector $u_{\mathcal{A}}$ to be the one that has unit-length and produces an equiangle-valued vector $A_{\mathcal{A}} 1_{\mathcal{A}}$ when it post-multiplied to $X_{\mathcal{A}}^T$. Thus, $u_{\mathcal{A}}$ should satisfy the following equation.

$$X_{\mathcal{A}}^T u_{\mathcal{A}} = A_{\mathcal{A}} 1_{\mathcal{A}} \tag{2.6.6}$$

This equation has an obvious solution, namely,

$$u_{\mathcal{A}} = X_{\mathcal{A}}(X_{\mathcal{A}}^T X_{\mathcal{A}})^{-1}(A_{\mathcal{A}} 1_{\mathcal{A}}) = A_{\mathcal{A}} X_{\mathcal{A}}(X_{\mathcal{A}}^T X_{\mathcal{A}})^{-1} 1_{\mathcal{A}} \tag{2.6.7}$$

Substituting (2.6.7) into (2.6.6) yields the right-hand side of (2.6.6). Furthermore, in order to make $u_{\mathcal{A}}$ have unit-length, the following should be satisfied and this yields the value of $A_{\mathcal{A}}$.

$$\begin{aligned} 1 &= ||u|| = A_{\mathcal{A}} ||X_{\mathcal{A}}(X_{\mathcal{A}}^T X_{\mathcal{A}})^{-1} 1_{\mathcal{A}}|| \\ \Rightarrow \quad 1 &= A_{\mathcal{A}}^2 ||X_{\mathcal{A}}(X_{\mathcal{A}}^T X_{\mathcal{A}})^{-1} 1_{\mathcal{A}}||^2 \\ &= A_{\mathcal{A}}^2 (1_{\mathcal{A}}^T (X_{\mathcal{A}}^T X_{\mathcal{A}})^{-1} 1_{\mathcal{A}}) \\ \Rightarrow \quad A_{\mathcal{A}}^2 &= (1_{\mathcal{A}}^T (X_{\mathcal{A}}^T X_{\mathcal{A}})^{-1} 1_{\mathcal{A}})^{-1} \\ A_{\mathcal{A}} &= (1_{\mathcal{A}}^T (X_{\mathcal{A}}^T X_{\mathcal{A}})^{-1} 1_{\mathcal{A}})^{-1/2} \end{aligned} \tag{2.6.8}$$

Next, the value of γ should be chose to yield that some variable, which is not included in $\mathcal{A}$, shows as much magnitude of $\hat{c}_j$ as the variables in $\mathcal{A}$. Note that, from the formula (2.6.4), variables in $\mathcal{A}$ are correlated with current residual in the magnitude of

$$\begin{aligned} c_j(\gamma) &= x_j'(y - \tilde{y}(\gamma)) \\ &= x_j'(y - \hat{y}_{\mathcal{A}}) - \gamma x_j' u_{\mathcal{A}} \\ &= \hat{c}_j - \gamma a_j \end{aligned} \tag{2.6.9}$$

$$\text{For } j \in \mathcal{A},\ |c_j(\gamma)| = \hat{C} - \gamma A_{\mathcal{A}} \tag{2.6.10}$$

where $a = X^T u_{\mathcal{A}}$. It should be noted that equation (2.6.10) represents that absolute values of correlations of variables in A are decreased in the same degree at each step.

Among the variables in the active set, the updating select the one producing the same absolute value of (2.6.9) with (2.6.10) using the smallest γ. Therefore,

for $j \in \mathcal{A}^c$,

$$|c_j(\gamma)| = |\hat{c}_j - \gamma a_j| = \hat{C} - \gamma A_{\mathcal{A}}$$

$$\text{i) } c_j(\gamma) \geq 0, \quad \hat{c}_j - \gamma a_j = \hat{C} - \gamma A_{\mathcal{A}} \;\Rightarrow\; \gamma = \frac{\hat{C} - \hat{c}_j}{A_{\mathcal{A}} - a_j}$$

$$\text{ii) } c_j(\gamma) < 0, -\hat{c}_j + \gamma a_j = \hat{C} - \gamma A_{\mathcal{A}} \;\Rightarrow\; \gamma = \frac{\hat{C} + \hat{c}_j}{A_{\mathcal{A}} + a_j}$$

Therefore, $\hat{\gamma}$ is determined as follows.

$$\hat{\gamma} = \min_{j \in \mathcal{A}}{}^{+}\left(\frac{\hat{C} - \hat{c}_j}{A_{\mathcal{A}} - a_j}, \frac{\hat{C} + \hat{c}_j}{A_{\mathcal{A}} + a_j}\right). \tag{2.6.11}$$

Here $\min^+$ denotes that only the positive components are considered in the minimization. And x_{jk}, the selected variable in the k-th step, is included in the active set.

Continuing this procedure, each step put one variable into the active set $\mathcal{A}$. Thus, its maximum number of steps is equal to the number of variables we are interested in. This efficiency is the most valuable strength of LARS. Note that this equiangular strategy doesn't seem to have any connection with Lass Regularization. However, Efron et al.(2004) point out that some modification of LARS produces the same result with the original Lasso estimation. Hence, LARS can be exploited as one of most efficient algorithm to compute Lasso solutions. Details on this point written in their paper may be worthwhile for the Interested readers.

The figure 2.5 represents an example of the LARS procedure on $p = 2$ case Efron et al.(2004) presented. Starting at the point $\hat{y}_0$, LARS estimate is updated by the equiangular strategy. In the figure, $\bar{y}_2$ denotes the projection of the response variable onto the column space of X. Thus, the response y affect current correlations of regressor variables only through $\bar{y}$. In this regard, x_1 has higher

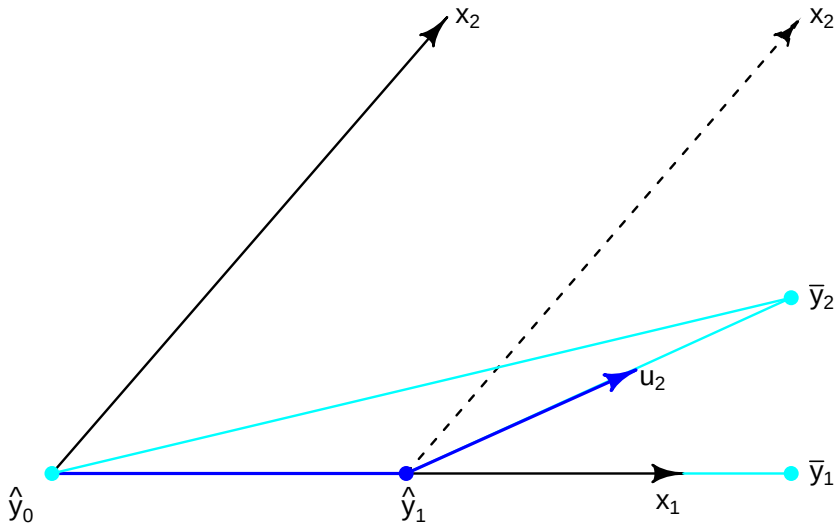

Figure 2.5: The Equiangular Strategy in LARS($p = 2$ case). The figure is from Figure 2. in Efron et al.(2004)

current correlation than x_2. Therefore, the first updating is carried out along the direction of x_1. Note that the result estimate is equal to $\hat{y}_1 = \hat{y}_0 + \hat{\gamma}x_1$, where is chosen so that x_1 and x_2 has the same current correlation with respect to the current residual. This can be seen that the angle between x_1 line and vector $\bar{y}_2 - \hat{y}_1$ are equal to the angle between dotted x_2 line and $\bar{y}_2 - \hat{y}_1$. Furthermore, the next step is carried out along the vector $\bar{y}_2 - \hat{y}_1$ since it equally divides the angle. Thus, equiangular vector u_2 is determined in the same direction of $\bar{y}_2 - \hat{y}_1$, but with unit-length. And the next estimate is computed as $\hat{y}_2 = \hat{y}_1 + \hat{\gamma}u_2$. Efron et al.(2004) mentioned that in $p = 2$ case, $\hat{y}_2$ becomes equal to $\bar{y}_2$. But this is not the general property in that $\hat{\gamma}$ is determined as having smaller value in $p > 2$ case.

2.6.2 MM-Algorithm

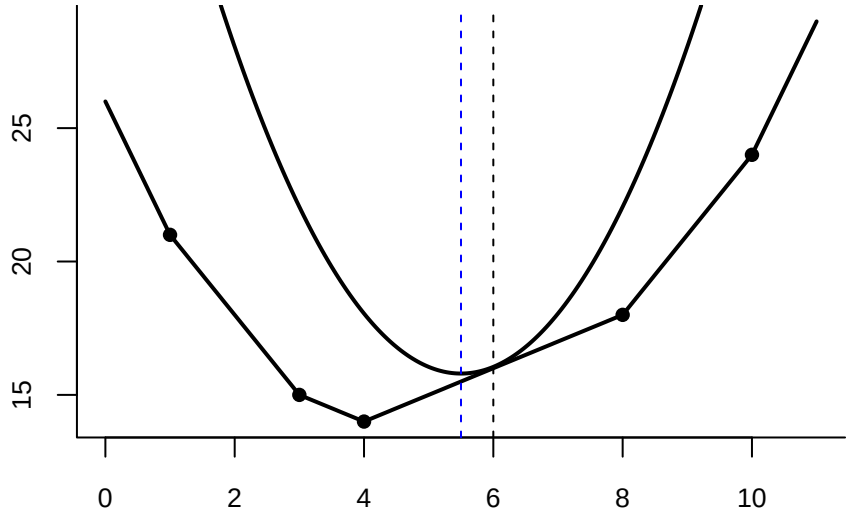

Figure 2.6: A quadratic majorizing function for the piecewise linear function. The figure is from Figure 8.1. in Lange(2013)

MM-algorithm(Hunter & Lange, 2004; Lange, 2013) is one of optimization algorithms that iteratively maximize or minimize a convex target function f using its surrogate function g. This surrogate function g is said to 'majorize' f in minimization problems, or 'minorize' it in maximization problems. Therefore this algorithm is a quite comprehensive for that it can be applied for both maximization and minimization problems. Without loss of generality, we discuss this algorithm in the case that the target function $f(\theta)$ is to be minimized.

When it is said that a function $g(\theta|\theta_{[m]})$ majorize $f(\theta)$ at $\theta_{[m]}$, this means the following two conditions are satisfied(Lange, 2013).

i. $f(\theta_{[m]}) = g(\theta_{[m]}|\theta_{[m]})$: g is tangent to f at $\theta = \theta_{[m]}$. (2.6.12)

ii. $f(\theta) \leq g(\theta|\theta_{[m]})$ for $\theta \neq \theta_{[m]}$: g is above f throughout the domain. (2.6.13)

Figure 2.6 shows one of the typical cases MM can be applied. The function f, which contains sum of absolute difference calculation, can not easily be dealt with in optimization problems. however, using its convexity and an appropriate surrogate function g, optimization problems can be solved with following steps.

1) Minimize $g(\theta|\theta_{[m]})$. And define its minimum point as $\theta_{[m+1]}$.
That is, $\theta_{[m+1]} = \underset{\theta}{\operatorname{argmin}}\big[g(\theta|\theta_{[m]})\big]$.
2) Find $g(\theta|\theta_{[m+1]})$, a new surrogate function of $f(\theta)$ at $\theta = \theta_{[m+1]}$.
3) Repeat step 1 and 2 iteratively, until the convergence is achieved.

In sum, by just minimizing g, minimization of f can also be achieved. The inequality below, which is called "The Descent Property" in Lange(2013), implies the reason why this algorithm works.

$$f(\theta_{[m+1]}) \leq g(\theta_{[m+1]} \mid \theta_{[m]}) \leq g(\theta_{[m]} \mid \theta_{[m]}) \leq f(\theta_{[m]}) \tag{2.6.14}$$

From this inequality, we obtain $f(\theta_{[m+1]}) \leq f(\theta_{[m]})$ and as $[m]$-steps move on, the target function f would be minimized. It should be noted that $\theta_{[m+1]}$ needs not to be the minimum point of $g(\theta|\theta_{[m]})$. The descent property holds if $g(\theta_{[m+1]}|\theta_{[m]}) \leq g(\theta_{[m]}|\theta_{[m]})$. Therefore this algorithm would be able to work even when the minimization of g is hard to achieve – it suffices to just decrease the function.

The most important thing in the MM-algorithm is how to find appropriate surrogate functions. Some of most widely used methods are described in Lange(2013). The present thesis introduce only a part of them, which are more useful for the extended application to SEM.

Definition of Convexity

$$f(\alpha_j t_j) \leq \sum_j \alpha_j f(t_j), \text{ where} \alpha_j > 0, \sum_j \alpha_j = 1, \text{ for } j = 1, \ldots, p. \tag{2.6.15}$$

Equation (2.6.15), which is well know as "Jensen's Inequality", is equal to the definition of "convex" function f. This can be used in finding a surrogate function in MM-algorithm. Let $t_j = \frac{x_j}{\alpha_j}(\theta_j - \theta_{[m]j}) + x^T\theta_{[m]}$, where x_j's are regressors in the model, θ is the parameter vector to be estimated, and $\theta_{[m]}$ is the estimates of θ in $[m]$-th step. α_j's can be defined freely as long as they satisfy the conditions in equation (2.6.15). One easy choice is $\alpha_j = \frac{|x_j|}{\sum_{k=1}^{p}|x_k|}$.

With these terms,

$$\begin{aligned} \textstyle\sum_j \alpha_j t_j &= \textstyle\sum_j (x_j\theta_j - x_j\theta_{[m]j}) + (\sum_j \alpha_j) x^T\theta_{[m]} \\ &= x^T\theta - x^T\theta_{[m]} + x^T\theta_{[m]} \\ &= x^T\theta \end{aligned} \tag{2.6.16}$$

Hence,

$$\begin{aligned} f(x^T\theta) = f(\textstyle\sum_j \alpha_j t_j) &\leq \textstyle\sum_j f(\frac{x_j}{\alpha_j}(\theta_j - \theta_{[m]j}) + x^T\theta_{[m]}) \\ &= g(\theta|\theta_m) \end{aligned} \tag{2.6.17}$$

Linearization

The second method can be used for a concave function f, which satisfies $f''(x) < 0$. Then, by the Taylor's theorem,

$$\begin{aligned} f(\theta_j) &= f(\theta_{[m]j}) + f'(\theta_{[m]j})(\theta_j - \theta_{[m]j}) + \frac{f''(\theta_j^*)}{2}(\theta_j - \theta_{[m]j}^*)^2 \\ &\leq f(\theta_{[m]j}) + f'(\theta_{[m]j})(\theta_j - \theta_{[m]j}) \\ &= g(\theta_j|\theta_{[m]j}) \end{aligned} \tag{2.6.18}$$

For example, let $f(\theta_j) = \sqrt{\theta_j}$. In this case,

$$\sqrt{\theta_j} \leq \sqrt{\theta_{[m]j}} + \frac{1}{2\sqrt{\theta_{[m]j}}}(\theta_j - \theta_{[m]j}) = g(\theta_j|\theta_{[m]j}) \tag{2.6.19}$$

Solving L_1-Regularization Problem using Lasso

For the end of this subsection, the application of MM-algorithm to Lasso-type regularization problem will be suggested as an example(Hunter & Lange, 2004; Hunter & Li, 2005; Lange, 2013; Paik, 2014). L_1-regularized loss function in the linear regression model was previously described in (2.3.11).

$$\begin{aligned} l(\beta) &= (y - X\beta)^T(y - X\beta) + \kappa \sum_{j=1}^{p} |\beta_j| \\ &= \sum_{i=1}^{n}(y_i - x_i\beta)^2 + \kappa \sum_{j=1}^{p} |\beta_j| \end{aligned}$$

As the first step of minimization of this function, its surrogate function will be derived. First, using the linearization method,

$$\begin{aligned} |\beta_j| = \sqrt{|\beta_j|^2} &\leq \sqrt{|\beta_{[m]j}|^2} + \frac{|\beta_j|^2 - |\beta_{[m]j}|^2}{2\sqrt{|\beta_{[m]j}|^2}} \\ &= \frac{2|\beta_{[m]j}|^2 + |\beta_j|^2 - |\beta_{[m]j}|^2}{2|\beta_{[m]j}|} \\ &= \frac{|\beta_j|^2}{2|\beta_{[m]j}|} + \frac{|\beta_{[m]j}|}{2} \end{aligned} \tag{2.6.20}$$

In addition, using the definition of convexity,

$$(y_i - x_i^T\beta)^2 \leq \sum_{j=1}^{p} \alpha_{ij}(y_i - \frac{x_{ij}}{\alpha_{ij}}(\beta_j - \beta_{[m]j}) - x_i\beta_{[m]}) \tag{2.6.21}$$

Finally, applying equation (2.6.20), (2.6.21) to L_1-regularized loss function (2.3.11),

$$\begin{aligned} l(\beta) &\leq \sum_{i=1}^{n}\sum_{j=1}^{p}\alpha_{ij}(y_i - \frac{x_{ij}}{\alpha_{ij}}(\beta_j - \beta_{[m]j}) - x_i\beta_{[m]})^2 + \kappa\sum_{j=1}^{p}(\frac{|\beta_j|^2}{2|\beta_{[m]j}|} + \frac{|\beta_{[m]j}|}{2}) \\ &= g(\beta|\beta_{[m]}) \qquad \text{where } \alpha_{ij} = \frac{|x_{ij}|}{\sum_{k=1}^{p}|x_{ik}|} \end{aligned} \tag{2.6.22}$$

According to the logic of MM-algorithm, the solution to the problem can be obtained by just minimizing $g(\beta|\beta_{[m]})$ in equation (2.6.22).

$$\hat{\beta}_{[m+1]j} = \hat{\beta}_j = \underset{\beta_j}{argmin}[g(\beta|\hat{\beta}_{[m]})] \tag{2.6.23}$$

Now, by differentiating equation (2.6.23) with respect to β_j,

$$\begin{aligned} &\frac{\partial g(\beta|\hat{\beta}_{[m]})}{\partial\beta_j} = -2\sum_{i=1}^{n}\alpha_{ij}\frac{x_{ij}}{\alpha_{ij}}(y_i - \frac{x_{ij}}{\alpha_{ij}}(\beta_j - \hat{\beta}_{[m]j}) - x_i\hat{\beta}_{[m]}) + \kappa\frac{\beta_j}{|\hat{\beta}_{[m]j}|} = 0 \\ &\Rightarrow -2\sum_{i=1}^{n}x_{ij}(y_i - x_i\hat{\beta}_{[m]}) + 2\sum_{i=1}^{n}\frac{x_{ij}^2}{\alpha_{ij}}(\beta_j - \hat{\beta}_{[m]j}) + \kappa\frac{\beta_j}{|\hat{\beta}_{[m]j}|} = 0 \\ &\Rightarrow (2\sum_{i=1}^{n}\frac{x_{ij}^2}{\alpha_{ij}} + \frac{\kappa}{|\hat{\beta}_{[m]j}|})\beta_j = 2\sum_{i=1}^{n}x_{ij}(y_i - x_i\hat{\beta}_{[m]}) + 2\sum_{i=1}^{n}\frac{x_{ij}^2}{\alpha_{ij}}\hat{\beta}_{[m]j} \\ &\Rightarrow \hat{\beta}_{[m+1]j} = \frac{2\sum_{i=1}^{n}x_{ij}(y_i - x_i\hat{\beta}_{[m]}) + 2\sum_{i=1}^{n}\frac{x_{ij}^2}{\alpha_{ij}}\hat{\beta}_{[m]j}}{2\sum_{i=1}^{n}\frac{x_{ij}^2}{\alpha_{ij}} + \frac{\kappa}{|\hat{\beta}_{[m]j}|}} \qquad j = 1, \ldots, p. \end{aligned} \tag{2.6.24}$$

Chapter 3
Bayesian Structural Equation Modeling

As described in the previous section, A Bayesian Approach can be understood as one of shrinkage estimation or regularization. Using this relationship, it is possible to apply several regularization methods to the linear regression model more easily. The present chapter expands this understanding to the structural equation modeling, which is the system of linear regression models. Most derivations can be found in previous researches on the issue(Song & Lee, 2012a; Guo et al., 2012). Some extending derivations, especially for Bayesian regularization of SEM, will be presented in Appendix C.

3.1 Basic Approach

One great advantage of the Bayesian analysis is that researchers can make use of the posterior distributions of the parameters they are interested in, unlike the classical statistical analysis where usually only the point and interval estimates are available. Researchers are able to make probabilistic decision making based on these distributions. Also when a new data set from the same phenomenon is obtained, Bayesian posterior distributions can easily be up-

dated as they are regarded as the prior, and combined with the new sample according to the similar calculation described above. This easy updating procedure, which is called 'Bayesian Updating', is another excellent advantage of Bayesian analysis.

Since SEM can be understood as the system of linear regressions with latent variables, Bayesian SEM can be attained by extending the results in section 2.6. There are several previous researches that studied on the Bayesian approaches to the factor analysis and structural equation modeling(Lee, 2007; Park & Casella, 2008; Song & Lee, 2012a,b; Guo et al., 2012; Wang, 2014). The next two subsection will deal with these issues.

3.1.1 Bayesian Factor Analysis

Bayesian Approach for factor analysis is well illustrated in Lee and Song's researches(Lee, 2007; Song & Lee, 2012a, b). They simply used conjugate priors for parameters in SEM so that their model was easily analyzed by means of Bayesian statistical methods.

In this chapter, only a brief summary on prior and posterior distributions is posted, with some changes in notation to make it fit with our description of SEM in (1.1.1)∼(1.1.4). In the following description, θ denotes the vector which contains all the parameters involved in the model. And y_i is a $p \times 1$ vector which contains all the observed variable values from i-th subject, as defined in Chapter 1. Additionally, y_j for $j = 1, \ldots, p$ denotes $n \times 1$ vector consisting of n values of j-th variable. Therefore, y_j^T is equal to j-th row of Y. And Λ_j denotes a column vector containing j-th row of Λ.

At first, Lee(2007) suggests the following priors for factor analysis model.

Prior Distributions for Bayesian FA

$$\cdot\ \omega_i|\ \theta \sim N_q(0, \Phi) \tag{3.1.1}$$

$$\cdot\ \psi_{\epsilon j}^{-1} \sim Gamma(\alpha_{0\epsilon j},\ \beta_{0\epsilon j}) \tag{3.1.2}$$

$$\cdot\ \Lambda_j|\ \psi_{\epsilon j} \sim N_q(\Lambda_{0j},\ \psi_{\epsilon j} H_{0yj}), \tag{3.1.3}$$

where H_{0yj} can be any positive definite matrix.

$$\cdot\ \Phi \sim IW_q(R_0^{-1},\ \rho_0) \tag{3.1.4}$$

where R_0 can be any positive definite matrix.

Here $\alpha_{0\epsilon j}, \beta_{0\epsilon j}, \Lambda_{0j}$ and ρ_0 are hyperparameters whose values are determined based on the previous researches, hypotheses, and so on. Combining these prior distribution and the likelihood function obtained from our model (1.1.1) $\sim$ (1.1.4), the following posterior distributions can be derived.

Posterior Distributions for Bayesian FA

$$\cdot\ \omega_i|\ y_i, \theta \sim N_q(\delta^T y_i,\ \Delta) \tag{3.1.5}$$

where $\delta = \Psi_\epsilon^{-1}\Lambda(\Lambda^T\Psi_\epsilon^{-1}\Lambda + \Phi^{-1})^{-1} = (\Lambda\Phi\Lambda^T + \Psi_\epsilon)^{-1}\Lambda\Phi$

$\Delta = (\Lambda^T\Psi_\epsilon^{-1}\Lambda + \Phi^{-1})^{-1} = \Phi - \Phi\Lambda^T(\Lambda\Phi\Lambda^T + \Psi_\epsilon)^{-1}\Lambda\Phi$

$$\cdot\ \psi_{\epsilon j}^{-1}|\ y_j, \Omega \sim Gamma(\alpha_{n\epsilon j}, \beta_{n\epsilon j}) \tag{3.1.6}$$

where $\alpha_{n\epsilon j} = \alpha_{0\epsilon j} + \dfrac{n}{2}$

$\beta_{n\epsilon j} = \beta_{0\epsilon j} + \dfrac{1}{2}(y_j^T y_j + \Lambda_{0j}^T H_{0yj}^{-1}\Lambda_{0j} - \Lambda_{nj}^T H_{nyj}^{-1}\Lambda_{nj})$

$$\cdot\ \Lambda_j|\ y_j, \Omega, \psi_{\epsilon j} \sim N_q(\Lambda_{nj},\ \psi_{\epsilon j} H_{nyj}), \tag{3.1.7}$$

where $\Lambda_{nj} = H_{nyj}(H_{0yj}^{-1}\Lambda_{0j} + \Omega y_j)$, $H_{nyj} = (H_{0yj}^{-1} + \Omega\Omega^T)^{-1}$

$$\cdot\ \Phi|\ Y, \Omega \sim IW_q(R_0^{-1} + \Omega\Omega^T,\ n + \rho_0) \tag{3.1.8}$$

3.1.2 Bayesian Structural Equation Modeling

The FA model described in the previous subsection is equal to the measurement model in the structural equation modeling. Therefore, expanding this model with the structural model part, Bayesian approach for SEM can be obtained. Some notational changes should be added. At first, note that $\Omega = (\eta^T, \xi^T)^T$ in SEM, unlike FA where $\Omega = \xi$. Therefore Φ in (3.1.1) and (3.1.5), which indicates the covariance matrix among ξ's, should be changed to Σ_ω in (1.1.9). Also the dimension of inverse wishart distribution in (3.1.4) and (3.1.8) is q_2, not $q = q_1 + q_2$ which equals to the number of all the latent variables. Additionally, the posterior distribution of Φ is now $IW_{q_2}(R_0^{-1} + \xi\xi^T,\ n + \rho_0)$.

Prior Distributions for Bayesian SEM

– Latent Scores

$$\cdot\ \omega_i|\ \theta \sim N_q(0, \Sigma_\omega) \tag{3.1.9}$$

– Measurement Model

$$\cdot\ \psi_{\epsilon j}^{-1} \sim Gamma(\alpha_{0\epsilon j},\ \beta_{0\epsilon j}) \tag{3.1.10}$$

$$\cdot\ \Lambda_j|\ \psi_{\epsilon j} \sim N_q(\Lambda_{0j},\ \psi_{\epsilon j} H_{0yj}), \tag{3.1.11}$$

where H_{0yj} can be any positive definite matrix.

– Structural Model

$$\cdot\ \psi_{\zeta k}^{-1} \sim Gamma(\alpha_{0\zeta k},\ \beta_{0\zeta k}) \tag{3.1.12}$$

$$\cdot\ \Lambda_{\eta k}|\ \psi_{\zeta k} \sim N_{q_{(k)}}(\Lambda_{0\eta k},\ \psi_{\zeta k} H_{0\eta k}), \tag{3.1.13}$$

where $H_{0\eta k}$ can be any positive definite matrix.

$$\cdot\ \Phi \sim IW_{q_2}(R_0^{-1},\ \rho_0) \tag{3.1.14}$$

In the above, $k = 1, \cdots, q$ is a subindex indicating the latent variables. Posterior distributions can be derived analogously to the previous cases.

Posterior Distributions for Bayesian SEM

– Latent Scores

$$\cdot\ \omega_i |\ y_i, \theta \sim N_q(\delta^T y_i,\ \Delta) \tag{3.1.15}$$

where $\delta = \Psi_\epsilon^{-1}\Lambda(\Lambda^T\Psi_\epsilon^{-1}\Lambda + \Sigma_\omega^{-1})^{-1} = (\Lambda\Sigma_\omega\Lambda^T + \Psi_\epsilon)^{-1}\Lambda\Sigma_\omega$

$\Delta = (\Lambda^T\Psi_\epsilon^{-1}\Lambda + \Sigma_\omega^{-1})^{-1} = \Sigma_\omega - \Sigma_\omega\Lambda^T(\Lambda\Sigma_\omega\Lambda^T + \Psi_\epsilon)^{-1}\Lambda\Sigma_\omega$

– Measurement Model

$$\cdot\ \psi_{\epsilon j}^{-1} |\ y_j, \Omega \sim Gamma(\alpha_{n\epsilon j}, \beta_{n\epsilon j}) \tag{3.1.16}$$

where $\alpha_{n\epsilon j} = \alpha_{0\epsilon j} + \dfrac{n}{2}$

$\beta_{n\epsilon j} = \beta_{0\epsilon j} + \dfrac{1}{2}(y_j^T y_j + \Lambda_{0j}^T H_{0yj}^{-1}\Lambda_{0j} - \Lambda_{nj}^T H_{nyj}^{-1}\Lambda_{nj})$

$$\cdot\ \Lambda_j |\ y_j, \Omega, \psi_{\epsilon j} \sim N_q(\Lambda_{nj},\ \psi_{\epsilon j}H_{nyj}), \tag{3.1.17}$$

where $\Lambda_{nj} = H_{nyj}(H_{0yj}^{-1}\Lambda_{0j} + \Omega y_j), \quad H_{nyj} = (H_{0yj}^{-1} + \Omega\Omega^T)^{-1}$

– Structural Model

$$\cdot\ \psi_{\zeta k}^{-1} |\ \Omega \sim Gamma(\alpha_{n\zeta k}, \beta_{n\zeta k}) \tag{3.1.18}$$

where $\alpha_{n\zeta k} = \alpha_{0\zeta k} + \dfrac{n}{2}$

$\beta_{n\zeta k} = \beta_{0\zeta k} + \dfrac{1}{2}(\eta_k^T\eta_k + \Lambda_{0\eta k}^T H_{0\eta k}^{-1}\Lambda_{0\eta k} - \Lambda_{n\eta k}^T H_{n\eta k}^{-1}\Lambda_{n\eta k})$

$$\cdot\ \Lambda_{\eta k} |\ \Omega, \psi_{\zeta k} \sim N_q(\Lambda_{n\eta k},\ \psi_{\zeta k}H_{n\eta k}), \tag{3.1.19}$$

where $\Lambda_{n\eta k} = H_{n\eta k}(H_{0\eta k}^{-1}\Lambda_{0\eta k} + \omega_{(k)}\eta_k),\ H_{n\eta k} = (H_{0\eta k}^{-1} + \omega_{(k)}\omega_{(k)}^T)^{-1}$

$$\cdot\ \Phi |\ Y, \Omega \sim IW_{q_2}(R_0^{-1} + \xi\xi^T,\ n + \rho_0) \tag{3.1.20}$$

3.2 Bayesian Regularization for SEM

A Bayesian Method for Regularizing SEM was first suggested by Guo et al.(2012). The authors implemented Lasso-type Regularization on SEM, modifying and applying prior distributions suggested in Park and Casella(2008) to the study on general Bayesian SEM in Lee(2007). However, their Lasso application was only on the structural model, not on the measurement model. In the present thesis, more general approach that regularizes both measurement and structural model will be suggested.

3.2.1 Bayesian Lasso for Factor Analysis

Bayesian Lasso for Factor Analysis model is a simple modification of Bayesian FA in (3.1.1) $\sim$ (3.1.8). The following is the priors for implementing BLasso to SEM that Guo et al.(2012) suggested. Note that priors for $\omega_i, \psi_{\epsilon j}^{-1}$, and Φ are equal to those in general Bayesian Factor Analysis.

Prior Distributions for BLasso FA

$$\cdot\ \omega_i|\ \theta \sim N_q(0, \Phi) \tag{3.2.1}$$

$$\cdot\ \psi_{\epsilon j}^{-1} \sim Gamma(\alpha_{0\epsilon j},\ \beta_{0\epsilon j}) \tag{3.2.2}$$

$$\cdot\ \Lambda_j|\ \psi_{\epsilon j}, \tau_{\Lambda_j} \sim N_{q_{(j)}}(\Lambda_{0j},\ \psi_{\epsilon j} H_{0yj}), \tag{3.2.3}$$

$$\text{where}\ \ H_{0yj} = diag(\tau_{\Lambda_j}),\ \tau_{\Lambda_j} = (\tau^2_{\Lambda_{j1}},\ \tau^2_{\Lambda_{j2}}, \ldots,\ \tau^2_{\Lambda_{jq_{(j)}}})^T$$

$$\cdot\ \pi(\tau_{\Lambda_j}|\ \kappa^2_{\Lambda_j}) \propto \prod_{k}^{q_{(j)}} \frac{\kappa^2_{\Lambda_j}}{2} \exp(-\kappa^2_{\Lambda_j} \tau^2_{\Lambda_{jk}}) \tag{3.2.4}$$

$$\cdot\ \kappa^2_{\Lambda_j} \sim Gamma(\alpha_{0\Lambda_j},\ \beta_{0\Lambda_j}) \tag{3.2.5}$$

$$\cdot\ \Phi \sim IW_q(R_0^{-1},\ \rho_0) \tag{3.2.6}$$

Here, $q_{(j)}$ is the number of latent regressors for j-th measurement. As in the general Bayesian SEM, the following posterior distributions can be derived from the priors above and the likelihood function corresponding to distributional assumptions on our data.

Posterior Distributions for BLasso FA

$$\cdot\ \omega_i|\ y_i, \theta \sim N_q(\delta^T y_i,\ \Delta) \tag{3.2.7}$$

where $\quad \delta = \Psi_\epsilon^{-1}\Lambda(\Lambda^T\Psi_\epsilon^{-1}\Lambda + \Phi^{-1})^{-1} = (\Lambda\Phi\Lambda^T + \Psi_\epsilon)^{-1}\Lambda\Phi$

$$\Delta = (\Lambda^T\Psi_\epsilon^{-1}\Lambda + \Phi^{-1})^{-1} = \Phi - \Phi\Lambda^T(\Lambda\Phi\Lambda^T + \Psi_\epsilon)^{-1}\Lambda\Phi$$

$$\cdot\ \psi_{\epsilon j}^{-1}|\ y_j, \Omega, \tau_{\Lambda_j} \sim Gamma(\alpha_{n\epsilon j}, \beta_{n\epsilon j}) \tag{3.2.8}$$

where $\ \alpha_{n\epsilon j} = \alpha_{0\epsilon j} + \dfrac{n}{2}$

$$\beta_{n\epsilon j} = \beta_{0\epsilon j} + \frac{1}{2}(y_j^T y_j + \Lambda_{0j}^T H_{0yj}^{-1}\Lambda_{0j} - \Lambda_{nj}^T H_{nyj}^{-1}\Lambda_{nj})$$

$$\cdot\ \Lambda_j|\ y_j, \Omega, \psi_{\epsilon j}, \tau_{\Lambda_j} \sim N_{q_{(j)}}(\Lambda_{nj},\ \psi_{\epsilon j}H_{nyj}), \tag{3.2.9}$$

where $\ \Lambda_{nj}\ = H_{nyj}(H_{0yj}^{-1}\Lambda_{0j} + \Omega y_j), \quad H_{nyj} = (H_{0yj}^{-1} + \Omega\Omega^T)^{-1}$

$$\cdot\ (1/\tau_{\Lambda_{jk}}^2)|\ \Lambda_{jk}, \psi_{\epsilon j}, \kappa_{\Lambda_j}^2 \sim IG(\mu_{\tau_{\Lambda_{jk}}}, \kappa_{\Lambda_j}^2),\ \ \mu_{\tau_{\Lambda_{jk}}} = \sqrt{\frac{\kappa_{\Lambda_j}^2}{(\lambda_{jk} - \lambda_{0jk})^2}\psi_{\epsilon j}} \tag{3.2.10}$$

$$\cdot\ \kappa_{\Lambda_j}^2|\ \tau_{\Lambda_j} \sim Gamma(\alpha_{n\Lambda_j}, \beta_{n\Lambda_j}) \tag{3.2.11}$$

where $\ \alpha_{n\Lambda_j} = \alpha_{0\Lambda_j} + q_{(j)}, \quad \beta_{n\Lambda_j} = \beta_{0\Lambda_j} + \dfrac{1}{2}\displaystyle\sum_k^{q_{(j)}} \tau_{\Lambda_{jk}}^2$

$$\cdot\ \Phi|\ Y, \Omega \sim IW_q(R_0^{-1} + \Omega\Omega^T,\ n + \rho_0) \tag{3.2.12}$$

In the above description, IG indicates 'Inverse Gaussian' distribution, whose density is as follows.

$$f(x; \mu, \kappa^2) = (\frac{\kappa^2}{2\pi x^3})^{\frac{1}{2}}\ \exp(\frac{-\kappa^2(x-\mu)^2}{2\mu^2 x}) \tag{3.2.13}$$

3.2.2 Bayesian Lasso for Structural Equation Modeling

This section extends BLasso FA to BLasso SEM. For this purpose, priors should be given for the structural model so that corresponding posteriors can be derived. Prior and posterior distributions for $\psi_{\epsilon j}$, Λ_j, τ_{Λ_j}, and $\kappa^2_{\Lambda_j}$ are given unchanged as in (3.2.2)$\sim$(3.2.5) and (3.2.8)$\sim$(3.2.11). Here, $q_{(k)}$ indicates the number of regressor for η_k.

Prior Distributions for BLasso SEM

$\cdot\ \omega_i|\ \theta \sim N_q(0, \Sigma_\omega)$ (3.2.14)

– Measurement Model

$\cdot\ \psi_{\epsilon j}^{-1} \sim Gamma(\alpha_{0\epsilon j},\ \beta_{0\epsilon j})$ (3.2.15)

$\cdot\ \Lambda_j|\ \psi_{\epsilon j}, \tau_{\Lambda_j} \sim N_{q_{(j)}}(\Lambda_{0j},\ \psi_{\epsilon j} H_{0yj})$, (3.2.16)

where $H_{0yj} = diag(\tau_{\Lambda_j}),\ \tau_{\Lambda_j} = (\tau^2_{\Lambda_{j1}},\ \tau^2_{\Lambda_{j2}}, \ldots,\ \tau^2_{\Lambda_{jq_{(j)}}})^T$

$\cdot\ \pi(\tau_{\Lambda_j}|\ \kappa^2_{\Lambda_j}) \propto \prod_k^{q_{(j)}} \frac{\kappa^2_{\Lambda_j}}{2} \exp(-\kappa^2_{\Lambda_j} \tau^2_{\Lambda_{jk}})$ (3.2.17)

$\cdot\ \kappa^2_{\Lambda_j} \sim Gamma(\alpha_{0\Lambda_j},\ \beta_{0\Lambda_j})$ (3.2.18)

– Structural Model

$\cdot\ \psi_{\zeta k}^{-1} \sim Gamma(\alpha_{0\zeta k},\ \beta_{0\zeta k})$ (3.2.19)

$\cdot\ \Lambda_{\eta k}|\ \psi_{\zeta k}, \tau_{\Lambda_{\eta k}} \sim N_{q_{(k)}}(\Lambda_{0\eta k},\ \psi_{\zeta k} H_{0\eta k})$, (3.2.20)

where $H_{0\eta k} = diag(\tau_{\Lambda_{\eta k}}),\ \tau_{\Lambda_{\eta k}} = (\tau^2_{\Lambda_{\eta k1}},\ \tau^2_{\Lambda_{\eta k2}}, \ldots,\ \tau^2_{\Lambda_{\eta k q_{(k)}}})^T$

$\cdot\ \pi(\tau_{\Lambda_{\eta k}}|\ \kappa^2_{\Lambda_{\eta k}}) \propto \prod_l^{q_{(k)}} \frac{\kappa^2_{\Lambda_{\eta k}}}{2} \exp(-\kappa^2_{\Lambda_{\eta k}} \tau^2_{\Lambda_{\eta k l}})$ (3.2.21)

$\cdot\ \kappa^2_{\Lambda_{\eta k}} \sim Gamma(\alpha_{0\Lambda_{\eta k}},\ \beta_{0\Lambda_{\eta k}})$ (3.2.22)

$\cdot\ \Phi \sim IW_{q_2}(R_0^{-1},\ \rho_0)$ (3.2.23)

Posterior Distributions for BLasso SEM

– Latent Scores

$$\cdot\ \omega_i |\ y_i, \theta \sim N_q(\delta^T y_i,\ \Delta) \tag{3.2.24}$$

$$\text{where} \quad \delta = \Psi_\epsilon^{-1}\Lambda(\Lambda^T\Psi_\epsilon^{-1}\Lambda + \Sigma_\omega^{-1})^{-1} = (\Lambda\Sigma_\omega\Lambda^T + \Psi_\epsilon)^{-1}\Lambda\Sigma_\omega$$

$$\Delta = (\Lambda^T\Psi_\epsilon^{-1}\Lambda + \Sigma_\omega^{-1})^{-1} = \Sigma_\omega - \Sigma_\omega\Lambda^T(\Lambda\Sigma_\omega\Lambda^T + \Psi_\epsilon)^{-1}\Lambda\Sigma_\omega$$

– Measurement Model

$$\cdot\ \psi_{\epsilon j}^{-1} |\ y_j, \Omega, \tau_{\Lambda_j} \sim Gamma(\alpha_{n\epsilon j}, \beta_{n\epsilon j}) \tag{3.2.25}$$

$$\text{where}\ \ \alpha_{n\epsilon j} = \alpha_{0\epsilon j} + \frac{n}{2}$$

$$\beta_{n\epsilon j} = \beta_{0\epsilon j} + \frac{1}{2}(y_j^T y_j + \Lambda_{0j}^T H_{0yj}^{-1}\Lambda_{0j} - \Lambda_{nj}^T H_{nyj}^{-1}\Lambda_{nj})$$

$$\cdot\ \Lambda_j |\ y_j, \Omega, \psi_{\epsilon j}, \tau_{\Lambda_j} \sim N_{q_{(j)}}(\Lambda_{nj},\ \psi_{\epsilon j}H_{nyj}), \tag{3.2.26}$$

$$\text{where}\ \ \Lambda_{nj}\ \ = H_{nyj}(H_{0yj}^{-1}\Lambda_{0j} + \Omega y_j), \quad H_{nyj} = (H_{0yj}^{-1} + \Omega\Omega^T)^{-1}$$

$$\cdot\ (1/\tau_{\Lambda_{jk}}^2) |\ \Lambda_{jk}, \psi_{\epsilon j}, \kappa_{\Lambda_j}^2 \sim IG(\mu_{\tau_{\Lambda_{jk}}}, \kappa_{\Lambda_j}^2),\ \ \mu_{\tau_{\Lambda_{jk}}} = \sqrt{\frac{\kappa_{\Lambda_j}^2}{(\lambda_{jk} - \lambda_{0jk})^2}\psi_{\epsilon j}} \tag{3.2.27}$$

$$\cdot\ \kappa_{\Lambda_j}^2 | \tau_{\Lambda_j} \sim Gamma(\alpha_{n\Lambda_j}, \beta_{n\Lambda_j}) \tag{3.2.28}$$

$$\text{where}\ \ \alpha_{n\Lambda_j} = \alpha_{0\Lambda_j} + q_{(j)}, \quad \beta_{n\Lambda_j} = \beta_{0\Lambda_j} + \frac{1}{2}\sum_k^{q_{(j)}} \tau_{\Lambda_{jk}}^2$$

(Continued)

– Structural Model

$$\cdot\ \psi_{\zeta k}^{-1} |\ \Omega, \tau_{\Lambda_{\eta k}} \sim Gamma(\alpha_{n\zeta k}, \beta_{n\zeta k}) \tag{3.2.29}$$

where $\alpha_{n\zeta k} = \alpha_{0\zeta k} + \dfrac{n}{2}$

$$\beta_{n\zeta k} = \beta_{0\zeta k} + \frac{1}{2}(\eta_k^T \eta_k + \Lambda_{0\eta k}^T H_{0\eta k}^{-1} \Lambda_{0\eta k} - \Lambda_{n\eta k}^T H_{n\eta k}^{-1} \Lambda_{n\eta k})$$

$$\cdot\ \Lambda_{\eta k} |\ \Omega, \psi_{\zeta k}, \tau_{\Lambda_{\eta k}} \sim N_{q_{(k)}}(\Lambda_{n\eta k},\ \psi_{\zeta k} H_{n\eta k}), \tag{3.2.30}$$

where $\Lambda_{n\eta k} = H_{n\eta k}(H_{0\eta k}^{-1} \Lambda_{0\eta k} + \omega_{(k)} \eta_k)$ (3.2.31)

$$H_{n\eta k} = (H_{0\eta k}^{-1} + \omega_{(k)} \omega_{(k)}^T)^{-1}$$

$$\cdot\ (1/\tau_{\Lambda_{\eta kl}}^2) |\ \Lambda_{\eta k}, \psi_{\zeta k}, \kappa_{\Lambda_{\eta k}}^2 \sim IG(\mu_{\tau_{\Lambda_{\eta kl}}}, \kappa_{\Lambda_{\eta k}}^2) \tag{3.2.32}$$

where $\mu_{\tau_{\Lambda_{\eta kl}}} = \sqrt{\dfrac{\kappa_{\Lambda_{\eta k}}^2}{(\lambda_{\eta kl} - \lambda_{\eta 0kl})^2} \psi_{\zeta k}}$

$$\cdot\ \kappa_{\Lambda_{\eta k}}^2 |\ \tau_{\Lambda_{\eta k}} \sim Gamma(\alpha_{n\Lambda_{\eta k}}, \beta_{n\Lambda_{\eta k}}) \tag{3.2.33}$$

where $\alpha_{n\Lambda_{\eta k}} = \alpha_{0\Lambda_{\eta k}} + q_{(k)}, \quad \beta_{n\Lambda_{\eta k}} = \beta_{0\Lambda_{\eta k}} + \dfrac{1}{2} \displaystyle\sum_{l}^{q_{(k)}} \tau_{\Lambda_{\eta kl}}^2$

$$\cdot\ \Phi |\ Y, \Omega \sim IW_{q_2}(R_0^{-1} + \xi\xi^T,\ n + \rho_0) \tag{3.2.34}$$

3.3 Limitation

As they proposed the Bayesian Lasso for SEM, Guo et al.(2012) concluded that BLasso performs very well in SEM. Supporting this argument, they represented some examples including an artificial simulation and real data analysis. However, the results they provided do not seem to be strong enough to fully

back up the contention.

At first, in the simulation example, they failed to consider various conditions on the parameters involved in their semiparametric SEM. Only a single value is given to each of the parameters as a true value. For example, they set all the non-fixed elements in Λ, Ψ_ϵ, and Ψ_ζ to be 0.36. And also true values of $\phi_{11}, \phi_{12}, \phi_{22}$ in matrix Φ are specified to (1, 0.25, 1) over the whole simulation. It is implausible to argue that these values fully cover the scope of those parameter values.

Comprehensive results can be attained from the simulation manipulating all of the parameters by one and one. However, this is extremely cumbersome in SEM that has a large number of parameters. As a matter of fact, there is no necessity for carrying it out. In SEM, What is the most important is a population covariance matrix which produces population or sample data used in simulation studies. Note that the underlying covariance structure we are interested in is implied in the population covariance matrix. Therefore, even if simulation studies do not need to vary each of all the parameters, some of them should be manipulated in order to cover a wide scope of values in the population covariance matrix. To give an example, Φ can be an interesting candidate since manipulating this matrix draws different population covariance matrices in different magnitudes. Other parameters are can also be operated to conduct the same effect. But simultaneous manipulation on a couple of parameters may offset the impact of each other. Therefore, conditions should be given under the careful scrutiny for the produced population covariance matrices to cover wider range of type and magnitude of the matrix. Guo and colleagues' result lacks consideration of this issue – they carried out their simulation studies with only single population covariance. Hence their result cannot be accepted as a general performance of BLasso without further studies

regarding wider conditions on the covariance matrix.

The second point we shall point out is that the previous researchers did not present any comparison between BLasso and other existing estimation methods such as the maximum likelihood principle. Without comparing the performances of the methods, their claim arguing that BLasso shows a superior performance in fitting SEM cannot be supported. For example, even though the RMS, an index they suggested to measure the performance of fitting a semiparametric SEM, seems quite great in their BLasso result, it can be the case that ML or other fitting methods yield better RMS value. Thus, the result they presented is not sufficient to underpin their assertion.

Furthermore, as described earlier in Park and Casella(2008), BLasso does not have any capacity of shrinking nuisance parameters completely to zero. Since Guo et al. suggested only the mean of 100 replicated estimates, we have no idea on whether their method was able to produce some zero coefficient or not. But the mean and standard error of those repeated result and Park and Casella's comment imply that BLasso is not able to make the same effect with original Lasso in SEM; it seems that what they con do is producing only the close-zero results for some parameters and actually this is what we observed in our own simulation of the present thesis. Also note that for the unneeded and unnecessary misspecified coefficients, ML can yield similar results. That is, even though ML cannot produce zero-coefficients, it yields estimates close to zero. Plus, other regularization methods such as Ridge has the same effect as well. Therefore, without variable deletion effect shown in Lasso regularization, it cannot be said that BLasso is superior methods for simultaneous estimation and model selection.

At last, the most critical point on BLasso should be discussed here agin. As we described in Section 2.4, even if zero-valued estimates can be obtained

by BLasso, it is invalid to consider these coefficients to be removed from the original model. In Bayesian perspective, parameters corresponding to these zero-estimates are said to have posterior distributions with zero mean, or median, or any other central tendency which can be used as a posterior point estimate. Note that this Bayesian model are not exactly equal to the model without those parameters. Therefore, it seems implausible that BLasso can have the same shrinkage effect with Lasso. Despite the fact that BLasso may estimate the misspecified parameters more close to zero than ML does, the previous researchers' claim stating that BLasso has the property of simultaneous estimation and model selection should be reconsidered.

In sum, even though the result of Guo et al. seems great, a distinction must be drawn between their method and Lasso. Despite the fact that Posterior distributions derived by BLasso may be useful in statistical inference, it seems the method cannot be in harmony with the spirit of Lasso. In this regard, we will be able to find great significance in implementing Lasso to SEM; it can produce the result that Guo et al. intended to realize.

Chapter 4

Implementing Lasso to Structural Equation Modeling

In this chapter, an Algorithm for implementing Lasso to SEM will be suggested. Since SEM can be understood as an extension of linear regression models, this might seem quite simple. However, fundamental differences between those two models including the existence of latent variables and the complex structure represented by the system of a number of equations, make it difficult to achieve our goal. The algorithm is obtained based on the maximum likelihood principle, EM-algorithm, regularization and several optimization methods such as LARS and MM to deal with this difficulty.

The derivation of the algorithm will be presented in the following order. First, likelihood functions related to the measurement model and structural model will be derived. These functions are used as a basic loss function during the whole derivation process. Next, L_1-regularization, namely Lasso estimation will be implemented to these functions. However, since these functions contains some unobservable variables, it cannot be minimized directly. To cope with this problem, "EM-Algorithm" will be introduced. In this algorithm, conditional expectations of log likelihood functions will be calculated and the penalty term for L_1-regularization will be added. Finally, optimization methods will be applied to minimize these conditional expectations. For this

phase, methods previously explained in chapter 2 - LARS and MM algorithms will be used.

4.1 Likelihood Functions in SEM

4.1.1 Measurement Model Part

From the model (1.1.3), (1.1.4) described in Chapter 1, we can obtain the likelihood function of the measurement model.

$$
\begin{aligned}
L_{Y,\Omega}(\Lambda, \Psi_\epsilon) &= \prod_{i=1}^{n} P(y_i, \omega_i) = \prod_{i=1}^{n} P(y_i|\omega_i)P(\omega_i) \qquad (4.1.1)\\
&= |2\pi\Psi_\epsilon|^{-\frac{n}{2}} \exp(-\frac{1}{2}\sum_{i=1}^{n}(y_i - \Lambda\omega_i)^T\Psi_\epsilon^{-1}(y_i - \Lambda\omega_i)) \times |2\pi\Sigma_\omega|^{-\frac{n}{2}} \exp(-\frac{1}{2}\sum_{i=1}^{n}{\omega_i}^T\Sigma_\omega^{-1}\omega_i)\\
&= |2\pi\Psi_\epsilon|^{-\frac{n}{2}} \exp(-\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{p}{\psi_\epsilon}_j^{-1}(y_{ji} - \Lambda_j^T\omega_i)^2) \times |2\pi\Sigma_\omega|^{-\frac{n}{2}} \exp(-\frac{1}{2}\sum_{i=1}^{n}{\omega_i}^T\Sigma_\omega^{-1}\omega_i)
\end{aligned}
$$

This result was already shown in Rubin and Thayer(1982) and Choi(2010) with some differences in notation. However, what they concerned with was the 'Exploratory Factor Analysis'(EFA) model. Since the factor loading matrix in the 'Confirmatory Factor Analysis'(CFA) model contains sparse structure, with some elements fixed to zero, the mathematical expression (1.1.3) can cover both of exploratory and confirmatory models; in the latter, unnecessary elements in ω_i's are left out automatically by those zeros. However, the method we suggest needs a distinction between equations of two types of factor analysis. We shall elaborate on this point first.

Let Λ_j be a $q \times 1$ column vector containing j-th row of Λ, and $\Lambda_{(j)}$ be another column vector which contains only the non-zero elements in Λ_j. As in Chapter 3, we denote the number of latent regressors for j-th measurement variable as

$q_{(j)}$, thus $\Lambda_{(j)}$ is a $q_{(j)} \times 1$ column vector. Also, let $\omega_{(j)i}$ be a vector including the latent scores corresponding to the elements in $\Lambda_{(j)}$; that is, it only contains a subset of ω_i. By the sparsity of Λ, in the CFA model, the j-th equation for i-th subject in the measurement model can be expressed as follows.

$$\begin{aligned} y_{ji} &= \Lambda_j^T \omega_i + \epsilon_{ji} \\ &= \Lambda_{(j)}^T \omega_{(j)i} + \epsilon_{ji} \end{aligned} \tag{4.1.2}$$

With this notation, the likelihood above can be re-expressed as follows.

$$\begin{aligned} & L_{Y,\Omega}(\Lambda, \Psi_\epsilon) \\ &= |2\pi\Psi_\epsilon|^{-\frac{n}{2}} \exp(-\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{p} {\psi_\epsilon}_j^{-1}(y_{ji} - \Lambda_j^T\omega_i)^2) \times |2\pi\Sigma_\omega|^{-\frac{n}{2}} \exp(-\frac{1}{2}\sum_{i=1}^{n} {\omega_i}^T\Sigma_\omega^{-1}\omega_i) \\ &= |2\pi\Psi_\epsilon|^{-\frac{n}{2}} \exp(-\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{p} {\psi_\epsilon}_j^{-1}(y_{ji} - \Lambda_{(j)}^T\omega_{(j)i})^2) \times |2\pi\Sigma_\omega|^{-\frac{n}{2}} \exp(-\frac{1}{2}\sum_{i=1}^{n} {\omega_i}^T\Sigma_\omega^{-1}\omega_i) \end{aligned} \tag{4.1.3}$$

Taking logarithms on both sides,

$$\begin{aligned} LL_M &\stackrel{\text{def}}{\equiv} \log(L_{Y,\Omega}(\Lambda, \Psi_\epsilon)) \\ &= -\frac{np}{2}\log(2\pi) - \frac{n}{2}\log(|\Psi_\epsilon|) - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{p} {\psi_\epsilon}_j^{-1}(y_{ji} - \Lambda_{(j)}^T\omega_{(j)i})^2 \\ &\quad - \frac{nq}{2}\log(2\pi) - \frac{n}{2}\log(|\Sigma_\omega|) - \frac{1}{2}\sum_{i=1}^{n} {\omega_i}^T\Sigma_\omega^{-1}\omega_i \\ &= -\frac{n}{2}\sum_{j=1}^{p}\log(\psi_{\epsilon j}) - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{p}\psi_{\epsilon j}^{-1}(y_{ji}^2 - 2y_{ji}\Lambda_{(j)}^T\omega_{(j)i} + \Lambda_{(j)}^T\omega_{(j)i}\omega_{(j)i}^T\Lambda_{(j)}) \\ &\quad - \frac{1}{2}\sum_{i=1}^{n} tr(\Sigma_\omega^{-1}\omega_i{\omega_i}^T) + const. \end{aligned} \tag{4.1.4}$$

4.1.2 Structural Model Part

Now, we represent the structural model (1.1.6) and (1.1.7) in a different way to derive the likelihood function in structural model. Let η_k be the k-th latent endogenous variable, $\omega_{(k)}$ be a matrix includes n responses' scores of $q_{(k)}$ regressors for η_k. Also let Λ_{η_k} be a row vector which includes path coefficients on $\omega_{(k)}$, and ζ_k be a corresponding error term for η_k. With these expressions, the k-th equation in the structural model is

$$\begin{gathered} \eta_k = \Lambda_{\eta_k}^T \omega_{(k)} + \zeta_k \\ k = 1, \ldots, q_1 \\ \eta_k : 1 \times n, \quad \omega_{(k)} : q_{(k)} \times n, \quad \Lambda_{\eta_k} : q_{(k)} \times 1, \quad \zeta_k : 1 \times n \end{gathered} \tag{4.1.5}$$

For the i-th response,

$$\begin{gathered} \eta_{ki} = \Lambda_{\eta_k}^T \omega_{(k)i} + \zeta_{ki} \\ i = 1, \ldots, n, \quad k = 1, \ldots, q_1 \\ \eta_{ki} : 1 \times 1, \quad \omega_{(k)i} : q_{(k)} \times 1, \quad \zeta_{ki} : 1 \times 1 \end{gathered} \tag{4.1.6}$$

By definition, $\omega_{(k)i} \sim N(0, \Sigma_{\omega(k)})$. where $\Sigma_{\omega(k)} = I_{q(\mathcal{A}_k{}^c)} \Sigma_\omega I_{q(\mathcal{A}_k{}^c)}^T$. Some undefined notations are used here. $I_{q(\mathcal{A}_k{}^c)}$ is a submatrix of q-dimension identity matrix I, with $\mathcal{A}_k{}^c$ rows deleted. When this matrix is pre-multiplied to a target matrix, it removes target's rows selected in sub-parentheses. And Its transpose, if post-multiplied, removes the corresponding columns of the target matrix. For example,

$$I_{3(\{2\})} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \qquad I_{4(\{1,3\})} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

$$I_{3(\{2\})} \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} I_{3(\{2\})}^T = \begin{pmatrix} a_{11} & a_{13} \\ a_{31} & a_{33} \end{pmatrix}$$

And $\mathscr{A}_k$ denotes the subset containing all sub-index numbers of regressors for η_k. Its complement, ${\mathscr{A}_k}^c$, contains the other variables. For example, assume that there are 3 latent endogenous variables $\{\eta_1, \eta_2, \eta_3\}$ and 2 latent exogenous variables $\{\xi_1, \xi_2\}$ in a research model. And let $\{\eta_1, \xi_1\}$ be regressors for η_2. In this case,

$$\begin{aligned} \Omega = \{\omega_1,\ \omega_1,\ \ldots,\ \omega_5\}^T &= \{\eta_1,\ \eta_2,\ \eta_3,\ \xi_1,\ \xi_2\}^T \\ \mathscr{A}_2 = \{1, 4\}, \quad {\mathscr{A}_2}^c &= \{2,\ 3,\ 5\} \\ \Sigma_{\omega(2)} = I_{5({A_2}^c)} \Sigma_\omega I_{5({A_2}^c)}^T &= I_{5(\{2,3,5\})} \Sigma_\omega I_{5(\{2,3,5\})}^T \\ = \begin{pmatrix} \Sigma_{\omega 11} & \Sigma_{\omega 14} \\ \Sigma_{\omega 41} & \Sigma_{\omega 44} \end{pmatrix} & \end{aligned}$$

From the above model, it can be known that given the corresponding regressor $\omega_{(k)i}$, η_{ki} follows the iid normal distribution with mean $\Lambda_{\eta_k}^T \omega_{(k)i}$ and variance ψ_{ζ_k}. However, there is an additional condition – recursiveness of the model required to derive the likelihood function of the structural model from its decomposition into the η_{ki}'s independent conditional distributions. The recursiveness indicates that the variables in SEM model have the only unidirectional causal relationships. In this case the model has no feedback path

between any two variables, whether the path is direct of indirect. And also there is no covariance path among the error terms(Bollen, 1989).

The following system implies the recursive relationship.

$$\begin{aligned}
\eta_1 &= \Gamma_1\xi + \zeta_1 \\
\eta_2 &= \beta_{21}\eta_1 + \Gamma_2\xi + \zeta_2 \\
\eta_3 &= \beta_{31}\eta_1 + \beta_{32}\eta_2 + \Gamma_3\xi + \zeta_3 \\
&\vdots \\
\eta_{q_1} &= \beta_{q_1 1}\eta_1 + \beta_{q_1 2}\eta_2 + \cdots + \beta_{q_1,q_1-1}\eta_{q_1-1} + \Gamma_{q_1}\xi + \zeta_{q_1-1}
\end{aligned} \tag{4.1.7}$$

where β_{ij} is (i,j)-th element in matrix B, Γ_i indicates the i-th row of Γ. Some path coefficient could be zero in the above equations.

In SEM, this recursiveness condition yields a low triangular matrix B and diagonal Ψ_ζ. Also, in this kind of model, the following relationships hold.

$$\omega^*_{(q_1)} \supseteq \omega^*_{(q_1-1)} \supseteq \cdots \supseteq \omega^*_{(1)} \tag{4.1.8}$$

$$\omega^*_{(k)} \supset \{\eta_{k-1}, \eta_{k-2}, \ldots, \eta_1\} \tag{4.1.9}$$

$$\omega^*_{(k)} = \{\eta_{k-1}, \omega^*_{(k-1)}\} \tag{4.1.10}$$

where $\omega^*_{(k)}$ represents a subset of all the possible regressors of η_k, including the elements in B and Γ. Note that this term is distinct from $\omega_{(k)}$ in (4.1.5) and (4.1.6), which indicates a subset of all the regressors actually included in the η_k's equation. Using those relationships, We can derive the $L(\Lambda_\eta, \Psi_\zeta)$, likelihood function of the structural model.

$$
\begin{aligned}
L(\Lambda_\eta, \Psi_\zeta) &= P(\eta_1, \eta_2, \ldots, \eta_{q_1}, \omega^*_{(1)}, \omega^*_{(2)}, \ldots, \omega^*_{(q_1)}) \\
&= P(\eta_{q_1}, \omega^*_{(q_1)}) \\
&= P(\eta_{q_1}|\omega^*_{(q_1)})\ P(\omega^*_{(q_1)}) \\
&= P(\eta_{q_1}|\omega^*_{(q_1)})\ P(\eta_{q_1-1}, \omega^*_{(q_1-1)}) \\
&= P(\eta_{q_1}|\omega^*_{(q_1)})\ P(\eta_{q_{1-1}}|\omega^*_{(q_1-1)}) P(\omega^*_{(q_1-1)}) \\
&= P(\eta_{q_1}|\omega^*_{(q_1)})\ P(\eta_{q_1-1}|\omega^*_{(q_1-1)}) P(\eta_{q_1-2}, \omega^*_{(q_1-2)}) \\
&= \cdots \\
&= P(\eta_{q_1}|\omega^*_{(q_1)})\ P(\eta_{q_1-1}|\omega^*_{(q_1-1)}) \cdots P(\eta_1|\omega^*_{(1)}) P(\omega^*_{(1)}) \\
&= P(\eta_{q_1}|\omega^*_{(q_1)})\ P(\eta_{q_1-1}|\omega^*_{(q_1-1)}) \cdots P(\eta_1|\omega^*_{(1)}) P(\xi) \\
&= P(\eta_{q_1}|\omega_{(q_1)})\ P(\eta_{q_1-1}|\omega_{(q_1-1)}) \cdots P(\eta_1|\omega_{(1)}) P(\xi)
\end{aligned}
\tag{4.1.11}
$$

Since $\xi \sim N(0, \Psi)$, $P(\xi)$ in the last line is constant with regard to $(\Lambda_\eta, \Psi_\zeta$, so it can be ignored. The above equations imply that in the recursive model, the structural model's likelihood can be obtained proportional to the product of η'_ks conditional density functions given the corresponding $\omega_{(k)}$'s, as follows.

$$
\begin{aligned}
L(\Lambda_\eta, \Psi_\zeta) &= \prod_{k=1}^{q_1} P(\eta_k|\omega_{(k)}) \times constant \\
&\propto \prod_{k=1}^{q_1} P(\eta_k|\omega_{(k)})
\end{aligned}
\tag{4.1.12}
$$

This make things simpler. Therefore we shall focus on the recursive model case henceforth. Using the result above, we can obtain log-likelihood function of the structural model.

$$
LL_S \stackrel{\text{def}}{\equiv} \log(L(\Lambda_\eta, \Psi_\zeta)) \propto \sum_{k=1}^{q_1} \log(P(\eta_k|\omega_{(k)})) \tag{4.1.13}
$$

The Log-likelihood function for η_k can be obtained as follows.

$$
\begin{aligned}
LL_{S_k} &\stackrel{\text{def}}{\equiv} \log(P(\eta_k|\omega_{(k)})) \qquad (4.1.14)\\
&= -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\psi_{\zeta k}) - \frac{1}{2}\sum_{i=1}^{n}\psi_{\zeta k}^{-1}(\eta_{ki} - \Lambda_{\eta k}^{T}\omega_{(k)i})^2\\
&= -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\psi_{\zeta k}) - \frac{1}{2}\sum_{i=1}^{n}\psi_{\zeta k}^{-1}(\eta_{ki}^2 - 2\eta_{ki}\omega_{(k)i}^{T}\Lambda_{\eta k} + \Lambda_{\eta k}^{T}\omega_{(k)i}\omega_{(k)i}^{T}\Lambda_{\eta k})
\end{aligned}
$$

Hence, by summing up LL_{S_k} for $k = 1, \ldots, q_1$, LL_S can be computed.

$$
\begin{aligned}
LL_S &= \sum_{k=1}^{q_1} LL_{S_k} + const. \qquad (4.1.15)\\
&= -\frac{n}{2}\sum_{k=1}^{q_1}\log(\psi_{\zeta k}) - \frac{1}{2}\sum_{k=1}^{q_1}\sum_{i=1}^{n}\psi_{\zeta k}^{-1}(\eta_{ki}^2 - 2\eta_{ki}\omega_{(k)i}^{T}\Lambda_{\eta k} + \Lambda_{\eta k}^{T}\omega_{(k)i}\omega_{(k)i}^{T}\Lambda_{\eta k})\\
&\quad + const.
\end{aligned}
$$

4.2 Double EM-algorithm for L_1-Regularized SEM

After deriving log-likelihood functions for both measurement and structural models, the next step is to optimize them so that we obtain our best estimates. However, as can be seen in the equations of those likelihoods, they contain unobservable latent variables. This makes it impracticable to deal with those functions.

A special technique, which is called 'EM-algorithm' can stand us in good stead regarding this situation. This algorithm is one of the most renowned method for computing maximum likelihood solution. Its strength lies in that it can be used for the case when there are some missing values in the current data. This might seem not that attractive, but careful scrutiny and creative intuition on defining 'missing' can maximize its potential versatility. For example, in our case we can consider the latent variables as missing. Then the EM-algorithm can be applied in an instant. In fact, this is what we shall carry out in this section.

Before moving on the application, a brief introduction to EM-algorithm will be illustrated. Consider the case we attempt to estimate parameter θ with data consisting of (Y, Z). Here Y represents observed data while Z refers to unobserved or unobservable data. The goal is to maximize the log-likelihood $L(\theta; Y)$, but note that

$$L(\theta; Y) = p(Y; \theta) = \int_Z p(Y, Z; \theta) dz \tag{4.2.1}$$

Since we don't have Z, the function is intractable and the maximization problem cannot be solved without any other alternatives.

EM-algorithm solves this problem by repeating two steps in turn. At first, let $LL(\theta; Y) = \log L(\theta; Y, Z)$ and $\hat{\theta}_{[m]}$ be our estimate in m-th iteration. In the

expectation step, which is abbreviated as E-step, conditional expectation of $LL(\theta; Y, Z)$ given Y and $\hat{\theta}_m$ is computed. We denote it as $Q(\theta; \theta_{[m]})$.

$$\begin{aligned} Q(\theta; \theta_{[m]}) &= E[LL(\theta; Y)|Y = y, \hat{\theta}_{[m]}] \\ &= \int_Z \log p(Y, Z; \theta) p(Z|Y, \hat{\theta}_{[m]}) dz \end{aligned} \tag{4.2.2}$$

Notice that the expectation is taken with respect to the conditional distribution of Z given Y and $\hat{\theta}_{[m]}$. Since we can derive this distribution, $Q(\theta; \theta_{[m]})$ is also manipulable.

In the following M-step, which stands for the maximization step, we obtain $\hat{\theta}_{[m+1]}$ as the value that maximizes $Q(\theta; \theta_{[m]})$. This step can be done with usual optimization methods since the preceding E-step makes the target function be tractable.

What is interesting is that we can reach the $\hat{\theta}^{MLE}$ that maximizes $L(\theta; Y)$ simply by repeating these two steps, which only deals with $Q(\theta; \theta_{[m]})$. Theorems proving the monotone increasing property of $L(\theta; Y)$ and the convergence of the estimate are well established in previous studies(Dempster, Laird & Rubin, 1977; Lange, 2013). Also note that we can still enjoy these properties even when we simply find $\hat{\theta}_{[m+1]}$ satisfying $Q(\theta_{[m+1]}; \theta_{[m]}) \geq Q(\theta_{[m]}; \theta_{[m]})$, not necessarily maximizing $Q(\theta; \theta_{[m]})$ in M-step. The replace version is known as 'Generalized EM-algorithm', or 'GEM algorithm'.

Some readers may aware of that the EM-algorithm resembles the MM-algorithm described in Chapter 2. As a matter of fact, EM is said to be a special case of MM, using $Q(\theta; \theta_{[m]})$ as the surrogate, or minorizing function of $LL(\theta; Y)$. Actually theorems regarding the excellent properties of EM are also related to those MM, such as descent property (4.2.26).

This section will illustrate the application of EM-algorithm to our problem.

However, since there are two models included in SEM and each has its own log likelihood function, the EM procedure should be conducted alternately; one for LL_M, the other for LL_S. Therefore "Double EM-Algorithm" that encompasses both of the log likelihood functions will be suggested to solve two optimization problems in SEM simultaneously.

4.2.1 E-step : Compute Conditional Expectations of Likelihood Functions

We start from LL_M, the **Measurement Model** log-likelihood function.

$$\begin{aligned} LL_M = &-\frac{n}{2}\sum_{j=1}^{p}\log(\psi_{\epsilon j}) - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{p}\psi_{\epsilon j}^{-1}(y_{ji}^2 - 2y_{ji}\Lambda_{(j)}^T\omega_{(j)i} + \Lambda_{(j)}^T\omega_{(j)i}\omega_{(j)i}^T\Lambda_{(j)}) \\ &- \frac{1}{2}\sum_{i=1}^{n} tr(\Sigma_\omega^{-1}\omega_i{\omega_i}^T) + const. \end{aligned}$$

In this function, we have all the y_i's, but the latent variable, or factor score ω_i's are not observed. Usually this function is called 'Complete Log-Likelihood' function since it contains all the variables involved. With EM-algorithm, we replace this function with 'Conditional Expectation of Log-Likelihood' function given the observed data and current estimates of parameters.

$$\begin{aligned} ELL_M &\stackrel{\text{def}}{\equiv} E(LL_M|Y,\theta) \\ &\propto -\frac{n}{2}\sum_{j=1}^{p}\log(\psi_{\epsilon j}) - \frac{1}{2}\sum_{i=1}^{n} tr(\Sigma_\omega^{-1}E(\omega_i{\omega_i}^T|Y,\theta)) \\ &\quad - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{p}\psi_{\epsilon j}^{-1}(y_{ji}^2 - 2y_{ji}\Lambda_{(j)}^T E(\omega_{(j)i}|Y,\theta) + \Lambda_{(j)}^T E(\omega_{(j)i}\omega_{(j)i}^T|Y,\theta)\Lambda_{(j)}) \end{aligned} \tag{4.2.3}$$

In order to obtain the conditional expectation of log-likelihood function, we

need $E(\omega_i {\omega_i}^T|Y,\theta)$, $E(\omega_{(j)i}|Y,\theta)$, and $E(\omega_{(j)i}{\omega_{(j)i}}^T|Y,\theta)$. Note that the last two terms are partitions of $E(\omega_i|Y,\theta)$ and $E(\omega_i{\omega_i}^T|Y,\theta)$, respectively. In order to obtain all of these terms, the conditional distribution of ω_i given Y and θ is required. This distribution is equal to equation (3.2.24) which is introduced as a part of posterior distributions in Bayesian/BLasso SEM.

$$\begin{aligned} \omega_i|y_i,\theta &\sim N(\delta^T y_i,\ \Delta) \\ \text{where } \delta &= (\Lambda\Sigma_\omega\Lambda^T+\Psi_\epsilon)^{-1}\Lambda\Sigma_\omega \ :\ p\times q \\ \Delta &= \Sigma_\omega - \Sigma_\omega\Lambda(\Lambda\Sigma_\omega\Lambda^T+\Psi_\epsilon)^{-1}\Lambda\Sigma_\omega \ :\ q\times q \end{aligned} \tag{4.2.4}$$

Hence, the conditional expectations can be obtained as follows.

$$\begin{aligned} &\cdot\ E(\omega_i|Y,\theta) = \delta^T y_i \\ &\cdot\ E(\omega_i{\omega_i}^T|Y,\theta) = Var(\omega_i|Y,\theta) + E(\omega_i|Y,\theta)E(\omega_i|Y,\theta)^T \\ &\qquad\qquad = \Delta + \delta^T y_i y_i^T\delta \\ &\cdot\ E(\omega_{(j)i}|Y,\theta) = \delta_{(j)}^T y_i \\ &\cdot\ E(\omega_{(j)i}{\omega_{(j)i}}^T|Y,\theta) = \Delta_{(j)} + \delta_{(j)}^T y_i y_i^T \delta_{(j)} \end{aligned} \tag{4.2.5}$$

Here $\delta_{(j)}$ indicates a $p\times q_{(j)}$ subset vector of δ, and $\Delta_{(j)}$ denotes a $q_{(j)}\times q_{(j)}$ partition of Δ, both of which contains the elements corresponding to latent regressor for j-th measurement equation. Furthermore, we introduce the following notations for convenience.

$$\begin{aligned} &\cdot\ W_i \overset{\text{def}}{\equiv} E(\omega_i{\omega_i}^T|Y,\theta) = \Delta + \delta^T y_i y_i^T\delta \\ &\cdot\ W \overset{\text{def}}{\equiv} \frac{1}{n}\sum_{i=1}^n W_i = \Delta + \delta^T\frac{YY^T}{n}\delta \\ &\cdot\ W_{(j)i} \overset{\text{def}}{\equiv} E(\omega_{(j)i}{\omega_{(j)i}}^T|Y,\theta) = \Delta_{(j)} + \delta_{(j)i}^T y_i y_i^T\delta_{(j)i} \\ &\cdot\ W_{(j)} \overset{\text{def}}{\equiv} \frac{1}{n}\sum_{i=1}^n W_{(j)i} = \Delta_{(j)} + \delta_{(j)}^T\frac{YY^T}{n}\delta_{(j)} \end{aligned} \tag{4.2.6}$$

Using these terms, the conditional expectation of log likelihood function becomes as in the following.

$$
\begin{aligned}
& ELL_M \qquad (4.2.7)\\
& \propto -\frac{n}{2}\sum_{j=1}^{p}\log(\psi_{\epsilon j}) - \frac{1}{2}tr(\Sigma_{\omega}^{-1}(nW)) - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{p}\psi_{\epsilon j}^{-1}(y_{ji}^2 - 2y_{ji}\Lambda_{(j)}^T\delta_{(j)}^T y_i + \Lambda_{(j)}^T(nW_{(j)})\Lambda_{(j)})\\
& \propto -\frac{n}{2}\sum_{j=1}^{p}\log(\psi_{\epsilon j}) - \frac{1}{2}tr(\Sigma_{\omega}^{-1}(nW)) - \frac{1}{2}\sum_{j=1}^{p}\psi_{\epsilon j}^{-1}(y_j^T y_j - 2\Lambda_{(j)}^T\delta_{(j)}^T Y y_j + \Lambda_{(j)}^T(nW_{(j)})\Lambda_{(j)})
\end{aligned}
$$

For the **Structural Model**, ELL_S can be defined and obtained analogously. The log likelihood for the structural model was

$$
\begin{aligned}
LL_S &= \sum_{k=1}^{q_1} LL_{S_k} + const. \qquad (4.2.8)\\
&= -\frac{n}{2}\sum_{k=1}^{q_1}\log(\psi_{\zeta k}) - \frac{1}{2}\sum_{k=1}^{q_1}\sum_{i=1}^{n}\psi_{\zeta k}^{-1}(\eta_{ki}^2 - 2\eta_{ki}\omega_{(k)i}^T\Lambda_{\eta k} + \Lambda_{\eta k}^T\omega_{(k)i}\omega_{(k)i}^T\Lambda_{\eta k})\\
&\quad + const.
\end{aligned}
$$

Taking conditional expectations,

$$
\begin{aligned}
& ELL_S \overset{\text{def}}{\equiv} E(LL_S|Y,\theta) \qquad (4.2.9)\\
& \propto -\frac{n}{2}\sum_{k=1}^{q_1}\log(\psi_{\zeta k})\\
& \quad - \frac{1}{2}\sum_{k=1}^{q_1}\sum_{i=1}^{n}\psi_{\zeta k}^{-1}(E(\eta_{ki}^2|Y,\theta) - 2E(\eta_{ki}\omega_{(k)i}^T,|Y,\theta)\Lambda_{\eta k} + \Lambda_{\eta k}^T E(\omega_{(k)i}\omega_{(k)i}^T|Y,\theta)\Lambda_{\eta k})
\end{aligned}
$$

To calculate this, three conditional expectation – 1) $E(\eta_{ki}^2|Y,\theta)$, 2) $E(\eta_{ki}\omega_{(k)i}^T,|Y,\theta)$, 3) $E(\omega_{(k)i}\omega_{(k)i}^T|Y,\theta)$ should be derived first. Actually, each of these terms is also a partition of the matrix W, defined in (4.2.6). W is a matrix containing the ex-

pected 2nd order moments among the latent variables. According to the definition of moment, each term above can be understood as

1) Conditional expectation of 2nd order moment of η_{ki}
2) Conditional expectation of 2nd order joint moments between η_{ki} and each element of $\omega_{(k)i}$
3) Conditional expectation of 2nd order (joint) moments among the elements of $\omega_{(k)i}$

To deal with these terms conveniently, some new notations are introduced again. First, let $W_{k,k}$ be a (k,k) element of W. This is k-th diagonal element which equals to expected 2nd order moment of η_{ki}. Next, let V_k be a subset of the transpose of the k-th row of W, whose elements are from the columns corresponding to $\mathcal{A}_k$. Therefore, V_k is a $q_{(k)} \times 1$ vector containing expected 2nd order joint moments between η_{ki} and each element of $\omega_{(k)i}$. Finally, let $W_{\mathcal{A}_k}$ denote the partition matrix of W, which contains only the rows and columns whose numbers are included in $\mathcal{A}_k$. This matrix is a $q_{(k)} \times q_{(k)}$ symmetric matrix that consists of expected 2nd order (joint) moments among the elements of $\omega_{(k)i}$.

These matrices and vector, $W_{k,k}$, V_k, and $W_{\mathcal{A}_k}$, indicate the above terms 1), 2), 3), respectively. As described earlier, All of these terms are related to the matrix W. So these terms can be expressed analogously as (4.2.6).

$$\begin{aligned} W_{i|k,k} &\stackrel{\text{def}}{\equiv} E(\eta_{ki}^2|Y,\theta) \;\; = \Delta_{k,k} + \delta_k y_i y_i^T \delta_k \\ W_{k,k} &\stackrel{\text{def}}{\equiv} \frac{1}{n}\sum_{i=1}^{n} W_{i|k,k} = \Delta_{k,k} + \delta_k \frac{YY^T}{n} \delta_k \end{aligned} \tag{4.2.10}$$

$$
\begin{aligned}
V_{i|\eta_k}^T &\stackrel{\text{def}}{\equiv} E(\eta_{ki}\omega_{(k)i}^T|Y,\theta) = \begin{pmatrix} \Delta_{k,l_1} + \delta_k^T y_i y_i^T \delta_{l_1} \\ \vdots \\ \Delta_{k,l_{q_k}} + \delta_k^T y_i y_i^T \delta_{l_{q_k}} \end{pmatrix} \\
V_{\eta_k}^T &\stackrel{\text{def}}{\equiv} \frac{1}{n}\sum_{i=1}^n V_{i|\eta_k}^T = \begin{pmatrix} \Delta_{k,l_1} + \delta_k^T \frac{YY^T}{n} \delta_{l_1} \\ \vdots \\ \Delta_{k,l_{q_k}} + \delta_k^T \frac{YY^T}{n} \delta_{l_{q_k}} \end{pmatrix}
\end{aligned}
\tag{4.2.11}
$$

$$
\begin{aligned}
W_{i|\mathcal{A}_k} &\stackrel{\text{def}}{\equiv} E(\omega_{(k)i}\omega_{(k)i}|Y,\theta) \\
&= I_{q(\mathcal{A}_k^c)} W_i I_{q(\mathcal{A}_k^c)}^T = \Delta_{\mathcal{A}_k} + \delta_{\mathcal{A}_k}^T y_i y_i^T \delta_{\mathcal{A}_k} \\
W_{\mathcal{A}_k} &\stackrel{\text{def}}{\equiv} \frac{1}{n}\sum_{i=1}^n W_{i|\mathcal{A}_k} \\
&= I_{q(\mathcal{A}_k^c)} W I_{q(\mathcal{A}_k^c)}^T = \Delta_{\mathcal{A}_k} + \delta_{\mathcal{A}_k}^T \frac{YY^T}{n} \delta_{\mathcal{A}_k}
\end{aligned}
\tag{4.2.12}
$$

Hence, with these matrices and vector, we can obtain ELL_S as in the following.

$$
ELL_S \propto -\frac{n}{2}\sum_{k=1}^{q_1}\log(\psi_{\zeta k}) - \frac{1}{2}\sum_{k=1}^{q_1}\psi_{\zeta k}^{-1}(nW_{k,k} - 2nV_{\eta_k}^T\Lambda_{\eta k} + \Lambda_{\eta k}^T(nW_{\mathcal{A}_k})\Lambda_{\eta k})
\tag{4.2.13}
$$

4.2.2 M-step : Minimizing the target function

In EM-algorithm, M-step stands for 'Maximization' or 'Minimization'. In this step, estimates would be obtained in the process of maximizing conditional expectation of log likelihood function. Often, -2 (conditional expectation of) log likelihood function is used as a target function, which is minimized during the estimation.

Before applying maximization or minimization methods to these functions, the penalty term should be added for regularization. $RELL$ functions, which stand for "Regularized Conditional Expectation of Log Likelihood Functions", are defined as follows.

$$RELL_M \stackrel{\text{def}}{\equiv} ELL_M - \frac{\kappa_m}{2} \sum_{j=1}^{p} \sum_{k}^{q_{(j)}} |\Lambda_{jk}| \tag{4.2.14}$$

$$RELL_S \stackrel{\text{def}}{\equiv} ELL_S - \frac{\kappa_s}{2} \sum_{k=1}^{q_1} \sum_{l}^{q_{(k)}} |\Lambda_{\eta kl}| \tag{4.2.15}$$

These functions play the role as the objective functions for L_1-regularized estimation of SEM. In penalty terms of the above functions, we denote the second summations as $\sum_{k}^{q_{(j)}}$ and $\sum_{l}^{q_{(k)}}$, instead of $\sum_{k=1}^{q_{(j)}}$ and $\sum_{l=1}^{q_{(k)}}$. By those subindices, we intend to mean that the summations only concern the slope coefficients corresponding to $q_{(j)}$ and $q_{(k)}$ latent regressors in the relevant equations, not all the $k = 1, 2, \ldots$ and $l = 1, 2, \ldots$ elements of the vector Λ_j and $\Lambda_{\eta k}$ in an ordinary order.

Measurement Model

$$1)\ \psi_{\epsilon j} : \hat{\psi}_{\epsilon j} = \underset{\psi_{\epsilon j}}{\text{argmin}}(-2RELL_M)$$

$$\frac{\partial(-2RELL_M)}{\partial\psi_{\epsilon j}} = \frac{n}{\psi_{\epsilon j}} + \frac{1}{\psi_{\epsilon j}^2}(y_j^T y_j - 2\Lambda_{(j)}^T\delta_{(j)}^T Y y_j + \Lambda_{(j)}^T(nW_{(j)})\Lambda_{(j)}) = 0$$

$$\Rightarrow \hat{\psi}_{\epsilon j} = \frac{1}{n}(y_j^T y_j - 2\Lambda_{(j)}^T\delta_{(j)}^T Y y_j + \Lambda_{(j)}^T(nW_{(j)})\Lambda_{(j)}) \tag{4.2.16}$$

$$2)\ \Lambda_j : \hat{\Lambda}_j \Leftarrow \hat{\Lambda}_{(j)} = \underset{\Lambda_{(j)}}{\text{argmin}}(-2RELL_M)$$

$$\begin{aligned}\hat{\Lambda}_{(j)} &= \underset{\Lambda_{(j)}}{\text{argmin}}(-2RELL_M) \\ &= \underset{\Lambda_{(j)}}{\text{argmin}}\Big(\sum_{j=1}^{p}\psi_{\epsilon j}^{-1}(y_j^T y_j - 2\Lambda_{(j)}^T\delta_{(j)}^T Y y_j + \Lambda_{(j)}^T(nW_{(j)})\Lambda_{(j)}) + \kappa_m\sum_{j=1}^{p}\sum_{k}^{q_{(j)}}|\Lambda_{jk}|\Big) \\ &= \underset{\Lambda_{(j)}}{\text{argmin}}\left(\frac{-2\Lambda_{(j)}^T\delta_{(j)}^T Y y_j}{\psi_{\epsilon j}} + \frac{\Lambda_{(j)}^T(nW_{(j)})\Lambda_{(j)}}{\psi_{\epsilon j}} + \kappa_m\sum_{k}^{q_{(j)}}|\Lambda_{jk}|\right)\end{aligned} \tag{4.2.17}$$

The minimization of (4.2.17) can be done by applying the 'LARS algorithm' or the 'MM-algorithm' described in Chapter 2. We shall elaborate on the details regarding this application later.

Structural Model

$$1)\ \psi_{\zeta k} : \hat{\psi}_{\zeta k} = \underset{\psi_{\zeta k}}{\text{argmin}}(-2RELL_S)$$

$$\frac{\partial(-2RELL_S)}{\partial\psi_{\zeta k}} = \frac{n}{\psi_{\zeta k}} + \frac{1}{\psi_{\zeta k}^2}(nW_{k,k} - 2nV_{\eta_k}^T\Lambda_{\eta k} + \Lambda_{\eta k}^T(nW_{\mathcal{A}_k})\Lambda_{\eta k}) = 0$$

$$\Rightarrow \hat{\psi}_{\zeta k} = W_{k,k} - 2V_{\eta_k}^T\Lambda_{\eta k} + \Lambda_{\eta k}^T W_{\mathcal{A}_k}\Lambda_{\eta k} \tag{4.2.18}$$

2) $\Lambda_{\eta k} : \hat{\Lambda}_{\eta k} = \underset{\Lambda_{\eta k}}{\text{argmin}}(-2RELL_S)$

$$\begin{aligned}\hat{\Lambda}_{\eta k} &= \underset{\Lambda_{\eta k}}{\text{argmin}}(-2RELL_S) \\ &= \underset{\Lambda_{\eta k}}{\text{argmin}}\Big(\sum_{k=1}^{q_1} \psi_{\zeta k}^{-1}(nW_{k,k} - 2nV_{\eta_k}^T\Lambda_{\eta k} + \Lambda_{\eta k}^T(nW_{\mathcal{A}_k})\Lambda_{\eta k}) + \kappa_s \sum_{k=1}^{q_1}\sum_{l}^{q_{(k)}} |\Lambda_{\eta k l}|\Big) \\ &= \underset{\Lambda_{\eta k}}{\text{argmin}}\Big(\frac{-2nV_{\eta_k}^T\Lambda_{\eta k}}{\psi_{\zeta k}} + \frac{\Lambda_{\eta k}^T(nW_{\mathcal{A}_k})\Lambda_{\eta k}}{\psi_{\zeta k}} + \kappa_s \sum_{l}^{q_{(k)}} |\Lambda_{\eta k l}|\Big)\end{aligned} \tag{4.2.19}$$

Minimization of (4.2.19) can be done analogously to that of (4.2.17). In the next section, some details will be discussed about applying LARS or MM to these minimization problems.

3) $\hat{\mathrm{P}}_{\omega}^{*}$: Temporary estimate for the factor correlation matrix.

i) $W^* = \frac{1}{2}[W + W^T]$, where W is obtained by (4.2.6).

ii) $\hat{\mathrm{P}}_{\omega}^{*} \leftarrow$ Standardize $W^*$by $diag(W^*)^{-1/2}\, W\, diag(W^*)^{-1/2}$ (4.2.20)

This estimate is proposed by Rubin and Thayer(1982) in their EM-algorithm approach to confirmatory factor analysis model. Some modification is added as in the step i), which is a suggestion by Adachi(2013) to guarantee the symmetry of the result.

However, unlike those previous studies that use this matrix as an EM-ML estimate for factor correlation matrix, P_{ω}^{*} cannot be treated as the same in SEM. This is because that the covariance matrix of latent variables, Σ_ω, is computed as a matrix-valued function of other parameters $B, \Gamma, \Psi_\zeta,$ and Φ as described in equation (1.1.10). Some may claim that the lower-right partition of P_{ω}^{*} corresponding to Φ in Σ_ω can be used as an estimate for Φ. This can be the case,

but when we optimize $RELL_S$ using LARS or MM, there is a further issue regarding the rescaling problem of the matrix P^*_ω. Therefore, we regard the result from the above equations only as the temporary estimate, and this will be used in achieving our final estimate for factor correlation matrix. However, bear in mind that only the partition corresponding to Φ can be used as our estimate. Estimates of the other parts in Σ_ω are obtained by calculating equation (1.1.10) using the estimates of B, Γ and Ψ_ζ. The rescaling issue will be discussed in the next section.

4.2.3 Optimization Methods for M-step

This section is supposed to introduce several issues on conducting the M-step in the previous section, using LARS or MM-algorithm. These algorithms can easily be applied to the M-step for L_1-regularized SEM, but some cautions are required before getting start.

1) LARS for L_1-Regularized SEM

As explained earlier in Chapter 2, LARS is one of the most useful algorithms for statistical analysts to fit the Lasso regularization to their data. However, some cautions on practical issues should be noted before we make use of this algorithm.

Most of computational programs with LARS algorithm requires their user to input regressor variable X, response variable y and a value of the tuning parameter κ. Then these programs conduct their estimation process, minimizing the following function.

$$l(\beta)\ (y - X\beta)^T(y - X\beta) + \kappa \sum_{j=1}^{p} |\beta_j| \tag{4.2.21}$$

Considering this fact, some pre-processes are needed to make these computational programs minimize the equations (4.2.17) and (4.2.19) exactly. The followings are modification and extension of Step (3.b) in Algorithm 1 in Choi(2010).

Measurement Model

Let $\tilde{y}$ and $\tilde{X}$ be response and regressor input to LARS program, respectively. These terms should be defined to satisfy the following equality, to make the program minimize equation (4.2.17).

$$\begin{aligned}
&\min_{\Lambda_{(j)}}\left[(\,\tilde{y}-\tilde{X}\Lambda_{(j)})^T(\tilde{y}-\tilde{X}\Lambda_{(j)})+\kappa_m\sum_{k}^{q_{(j)}}|\Lambda_{jk}|\right] \\
&\quad=-2\tilde{y}^T\tilde{X}\Lambda_{(j)}+\Lambda_{(j)}^T\tilde{X}^T\tilde{X}\Lambda_{(j)}+\kappa_m\sum_{k}^{q_{(j)}}|\Lambda_{jk}| \\
&\quad\equiv \min_{\Lambda_{(j)}}[-2RELL_M] \;=\; \frac{-2\Lambda_{(j)}^T\delta_{(j)}^T Y y_j}{\psi_{\epsilon j}}+\frac{\Lambda_{(j)}^T(nW_{(j)})\Lambda_{(j)}}{\psi_{\epsilon j}}+\kappa_m\sum_{k}^{q_{(j)}}|\Lambda_{jk}| \qquad (4.2.22)
\end{aligned}$$

$$\begin{aligned}
&\Rightarrow \quad \tilde{X}^T\tilde{X}=\frac{nW_{(j)}}{\psi_{\epsilon j}}, \qquad \tilde{y}^T\tilde{X}=\frac{y_jY^T\delta_{(j)}}{\psi_{\epsilon j}} \\
&\Rightarrow \quad \tilde{X}=chol(\frac{nW_{(j)}}{\psi_{\epsilon j}}), \qquad \tilde{y}=(\frac{y_jY^T\delta_{(j)}}{\psi_{\epsilon j}}\tilde{X}^{-1})^T
\end{aligned} \qquad (4.2.23)$$

'*chol*' indicates the 'Cholesky Decomposition'. However other matrix decomposition, such as 'Singular Value Decomposition' can be used. These two decomposition generate the exactly same results. Using these $\tilde{y}$ and $\tilde{X}$, the minimization problem (4.2.17) could be solved by LARS programs.

Structural Model

The logic described with respect to the measurement model can also be applied to the structural model. Let $\tilde{\eta}$ and $\tilde{\Omega}$ be the response and regressor input to LARS program, for the structural model. These terms should be defined

analogously to $\tilde{y}$ and $\tilde{X}$, but now the aim is to make the program minimize equation (4.2.19).

$$\begin{aligned}
&\min_{\Lambda_{\eta k}} \Big[(\tilde{\eta} - \tilde{\Omega}\Lambda_{\eta k})^T(\tilde{\eta} - \tilde{\Omega}\Lambda_{\eta k}) + \kappa_s \sum_{l}^{q_{(k)}} |\Lambda_{\eta k l}| \Big] \\
&\quad = -2\tilde{\eta}^T\tilde{\Omega}\Lambda_{\eta k} + \Lambda_{\eta k}^T\tilde{\Omega}^T\tilde{\Omega}\Lambda_{\eta k} + \kappa_s \sum_{l}^{q_{(k)}} |\Lambda_{\eta k l}| \\
&\quad \equiv \min_{\Lambda_{\eta k}}[-2RELL_S] \;=\; \frac{-2nV_{\eta k}^T\Lambda_{\eta k}}{\psi_{\zeta k}} + \frac{\Lambda_{\eta k}^T(nW_{\mathscr{A}_k})\Lambda_{\eta k}}{\psi_{\zeta k}} + \kappa_s \sum_{l}^{q_{(k)}} |\Lambda_{\eta k l}|
\end{aligned} \tag{4.2.24}$$

$$\begin{aligned}
&\Rightarrow \quad \tilde{\Omega}^T\tilde{\Omega} = \frac{nW_{\mathscr{A}_k}}{\psi_{\zeta k}}, \qquad \tilde{\eta}^T\tilde{\Omega} = \frac{nV_{\eta_k}\Lambda_{\eta k}}{\psi_{\zeta k}} \\
&\Rightarrow \quad \tilde{\Omega} = chol(\frac{nW_{\mathscr{A}_k}}{\psi_{\zeta k}}), \qquad \tilde{\eta} = (\frac{nV_{\eta_k}\Lambda_{\eta k}}{\psi_{\zeta k}}\tilde{\Omega}^{-1})^T
\end{aligned} \tag{4.2.25}$$

2) MM-algorithm for L_1-Regularized SEM

When it comes to the case we exploit the MM-algorithm in optimizing the $RELL$ functions, there are two strategies for this.

i) Apply the algorithm to $-2RELL$. That is, derive a surrogate function of $-2RELL$ and minimize it.

ii) Using the monotonicity of the expectation, the following inequality can be obtained immediately from equation (2.6.14).

$$E(f(x_m)) \leq E(g(x_{m+1}|x_m)) \leq E(g(x_m|x_m)) = E(f(x_m)) \tag{4.2.26}$$

The above inequality implies that the descent property also holds for the expectation of functions. Therefore MM-algorithm can be applied to regularized log likelihood function, before the expectation is taken. In this way, a surrogate function of regularized log likelihood function are derived first, and the

expectation of this surrogate will be regarded as the target function of the minimization problem.

In the present subsection, the application of MM-algorithm will be described using the second method since its result is more clear in representing equations of each coefficient. At first regularized log likelihood function should be obtained. This can be done by regularizing equations (4.1.4), (4.1.15) directly.

$$RLL_M \stackrel{\text{def}}{\equiv} LL_M - \frac{\kappa_m}{2}\sum_{j=1}^{p}\sum_{k}^{q_{(j)}}|\Lambda_{jk}| \tag{4.2.27}$$

$$RLL_S \stackrel{\text{def}}{\equiv} LL_S - \frac{\kappa_s}{2}\sum_{k=1}^{q_1}\sum_{l}^{q_{(}k)}|\Lambda_{\eta kl}| \tag{4.2.28}$$

From these definitions,

$$-2RLL_M \propto \sum_{i=1}^{n}\sum_{j=1}^{p}\psi_{\epsilon j}^{-1}(y_{ij}-\Lambda_j^T\omega_i)^2 + \kappa_m\sum_{j=1}^{p}\sum_{k}^{q_{(j)}}|\Lambda_{jk}| \tag{4.2.29}$$

$$-2RLL_S \propto \sum_{i=1}^{n}\sum_{k=1}^{q_1}\psi_{\zeta k}^{-1}(\eta_{ki}-\Lambda_{\eta k}^T\omega_{(k)i})^2 + \kappa_s\sum_{k=1}^{q_1}\sum_{l}^{q_k}|\Lambda_{\eta kl}| \tag{4.2.30}$$

In the following minimization processes, these functions are regarded as objective functions, and its surrogate functions will be derived. Then conditional expectations will be taken to these surrogates, and the result functions will be minimized.

Measurement Model

$$\begin{aligned} f_m(\Lambda) &= \sum_{i=1}^{n}\sum_{j=1}^{p}\psi_{\epsilon j}^{-1}(y_{ij}-\Lambda_j^T\omega_i)^2 + \kappa_m\sum_{j=1}^{p}\sum_{k}^{q_{(j)}}|\Lambda_{jk}| \qquad (4.2.31)\\ &\le \sum_{j=1}^{p}\sum_{i=1}^{n}\sum_{k}^{q_{(j)}}\frac{\alpha_{ki}}{\psi_{\epsilon j}}(y_{ji}-\frac{\omega_{ki}}{\alpha_{ki}}(\Lambda_{jk}-\Lambda_{[m]jk})-\omega_i^T\Lambda_{[m]j})^2 \\ &\quad + \kappa_m\sum_{j=1}^{p}\sum_{k}^{q_{(j)}}\left[\frac{\Lambda_{jk}^2}{2|\Lambda_{[m]jk}|}+\frac{|\Lambda_{[m]jk}|}{2}\right] \;=\; g_m(\Lambda|\Lambda_{[m]}) \end{aligned}$$

Note that with MM-algorithm, we do not need to lean on the second equation in (4.1.2), since MM decomposes the $q_{(j)}$ equations in the process of finding the surrogate function. In order to derive $E(g(\Lambda|\Lambda_{[m]})|Y,\theta)$, the conditional expectation for the term $(y_{ji} - \frac{\omega_{ki}}{\alpha_{ki}}(\Lambda_{jk} - \Lambda_{[m]jk}) - \omega_i^T\Lambda_{[m]j})^2$ should be computed first. Expanding the square,

$$
\begin{aligned}
&(y_{ji} - \frac{\omega_{ki}}{\alpha_{ki}}(\Lambda_{jk} - \Lambda_{[m]jk}) - \omega_i^T\Lambda_{[m]j})^2 \\
&= y_{ji}^2 + \frac{\omega_{ki}^2}{\alpha_{ki}^2}(\Lambda_{jk} - \Lambda_{[m]jk})^2 + \Lambda_{[m]j}^T\omega_i\omega_i^T\Lambda_{[m]j} - 2y_{ji}\omega_i^T\Lambda_{[m]j} \\
&\qquad - 2y_{ji}\frac{\omega_{ki}}{\alpha_{ki}}(\Lambda_{jk} - \Lambda_{[m]jk}) + 2\frac{\omega_{ki}}{\alpha_{ki}}(\Lambda_{jk} - \Lambda_{[m]jk})\omega_i^T\Lambda_{[m]j}
\end{aligned}
$$

Here, the first, third, and fourth terms can be omitted since they are constants with respect to Λ_j. And conditional expectations of remaining terms need $E(\omega_{ki}|Y,\theta)$, $E(\omega_{ki}^2|Y,\theta)$, $E(\omega_i|Y,\theta)$, $E(\omega_i\omega_i^T|Y,\theta)$, and $E(\omega_{ki}\omega_i^T|Y,\theta)$. Most of these terms can be obtained from equations (4.2.5), (4.2.10)$\sim$(4.2.12), except $E(\omega_{ki}\omega_i^T|Y,\theta)$, which can be expressed using the following definitions.

$$
\begin{aligned}
V_{i|k}^T \overset{\text{def}}{\equiv} E(\omega_{ki}\omega_i^T|Y,\theta) &= \begin{pmatrix} \Delta_{k,1} + \delta_k^T y_i y_i^T \delta_1 \\ \vdots \\ \Delta_{k,q} + \delta_k^T y_i y_i^T \delta_q \end{pmatrix} \\
V_k^T \overset{\text{def}}{\equiv} \frac{1}{n}\sum_{i=1}^n V_{i|k}^T &= \begin{pmatrix} \Delta_{k,1} + \delta_k^T \frac{YY^T}{n} \delta_1 \\ \vdots \\ \Delta_{k,q} + \delta_k^T \frac{YY^T}{n} \delta_q \end{pmatrix}
\end{aligned}
\tag{4.2.32}
$$

Therefore,

$$
\begin{aligned}
&E((y_{ji} - \frac{\omega_{ki}}{\alpha_{ki}}(\Lambda_{jk} - \Lambda_{[m]jk}) - \omega_i^T\Lambda_{[m]j})^2|Y,\theta) \\
&= y_{ji}^2 + \frac{E(\omega_{ki}^2|Y,\theta)}{\alpha_{ki}^2}(\Lambda_{jk} - \Lambda_{[m]jk})^2 + \Lambda_{[m]j}^T E(\omega_i\omega_i^T|Y,\theta)\Lambda_{[m]j} - 2y_{ji}E(\omega_i|Y,\theta)^T\Lambda_{[m]j} \\
&\quad - 2y_{ji}\frac{E(\omega_{ki}|Y,\theta)}{\alpha_{ki}}(\Lambda_{jk} - \Lambda_{[m]jk}) + 2\frac{1}{\alpha_{ki}}(\Lambda_{jk} - \Lambda_{[m]jk})(E(\omega_{ki}\omega_i^T|Y,\theta)\Lambda_{[m]j})
\end{aligned}
$$

$$= y_{ji}^2 + \frac{W_{i|k,k}}{\alpha_{ki}^2}(\Lambda_{jk} - \Lambda_{[m]jk})^2 + \Lambda_{[m]j}^T W_i \Lambda_{[m]j} - 2y_{ji} y_i^T \delta \Lambda_{[m]j} - 2y_{ji}\frac{\delta_k^T y_i}{\alpha_{ki}}(\Lambda_{jk} - \Lambda_{[m]jk})$$
$$+ 2\frac{1}{\alpha_{ki}}(\Lambda_{jk} - \Lambda_{[m]jk})(V_{i|k}^T \Lambda_{[m]j}) \tag{4.2.33}$$

$$E(g_m(\Lambda|\Lambda_{[m]})|Y,\theta)$$
$$= \sum_{j=1}^{p}\sum_{i=1}^{n}\sum_{k}^{q_{(j)}} \frac{\alpha_{ki}}{\psi_{\epsilon j}}\Big[y_{ji}^2 + \frac{W_{i|k,k}}{\alpha_{ki}^2}(\Lambda_{jk} - \Lambda_{[m]jk})^2 + \Lambda_{[m]j}^T W_i \Lambda_{[m]j} - 2y_{ji} y_i^T \delta \Lambda_{[m]j}$$
$$- 2y_{ji}\frac{\delta_k^T y_i}{\alpha_{ki}}(\Lambda_{jk} - \Lambda_{[m]jk}) + 2\frac{1}{\alpha_{ki}}(\Lambda_{jk} - \Lambda_{[m]jk})(V_{i|k}^T \Lambda_{[m]j})\Big]$$
$$+ \kappa_m \sum_{j=1}^{p}\sum_{k}^{q_{(j)}}\left[\frac{\Lambda_{jk}^2}{2|\Lambda_{[m]jk}|} + \frac{|\Lambda_{[m]jk}|}{2}\right] \tag{4.2.34}$$

Hence, by the MM-algorithm, our solution can be obtained by minimizing the function (4.2.34).

$$\frac{\partial E(g_m(\Lambda|\Lambda_{[m]})|Y,\theta)}{\partial \Lambda_{jk}}$$
$$= \sum_{i=1}^{n}\frac{\alpha_{ki}}{\psi_{\epsilon j}}\left[\frac{2W_{i|k,k}}{\alpha_{ki}^2}(\Lambda_{jk} - \Lambda_{[m]jk}) - 2y_{ji}\frac{\delta_k^T y_i}{\alpha_{ki}} + \frac{2}{\alpha_{ki}}(V_{i|k}^T \Lambda_{[m]j})\right] + \kappa_m \frac{\Lambda_{jk}}{|\Lambda_{[m]jk}|} = 0$$
$$\Rightarrow \Lambda_{jk}\left[2\sum_{i=1}^{n}\frac{\alpha_{ki}}{\psi_{\epsilon j}}\frac{W_{i|k,k}}{\alpha_{ki}^2} + \frac{\kappa_m}{|\lambda_{[m]jk}|}\right] = 2\sum_{i=1}^{n}\frac{\alpha_{ki}}{\psi_{\epsilon j}}\left[\frac{W_{i|k,k}}{\alpha_{ki}^2}\Lambda_{[m]jk} + y_{ji}\frac{\delta_k^T y_i}{\alpha_{ki}} - \frac{V_{i|k}^T \Lambda_{[m]j}}{\alpha_{ki}}\right]$$

$$\Rightarrow \hat{\Lambda}_{[m+1]jk} = \hat{\Lambda}_{jk} = \frac{\sum_{i=1}^{n}\left[\frac{W_{i|k,k}}{\alpha_{ki}}\Lambda_{[m]jk} + y_{ji}\delta_k^T y_i - V_{i|k}^T \Lambda_{[m]j}\right]}{\sum_{i=1}^{n}\frac{W_{i|k,k}}{\alpha_{ki}} + \frac{\kappa_m}{2|\lambda_{[m]jk}|}\psi_{\epsilon j}} \tag{4.2.35}$$

Structural Model

An MM-solution for the structural model can be obtained analogously. At first, we derive the surrogate function $g_s(\Lambda_\eta|\Lambda_{\eta[m]})$ as follows.

$$
\begin{aligned}
f_s(\Lambda_\eta) &= \sum_{i=1}^{n}\sum_{k=1}^{q_1}\psi_{\zeta k}^{-1}(\eta_{ki}-\Lambda_{\eta k}^T\omega_{(k)i})^2+\kappa_s\sum_{k=1}^{q_1}\sum_{l}^{q_{(k)}}|\Lambda_{\eta kl}| \qquad (4.2.36)\\
&\leq \sum_{i=1}^{n}\sum_{j=1}^{q_1}\sum_{l}^{q_{(k)}}\frac{\alpha_{li}}{\psi_{\zeta k}}(\eta_{ki}-\frac{\omega_{(k)li}}{\alpha_{li}}(\Lambda_{\eta kl}-\Lambda_{\eta[m]kl})-\omega_{(k)i}^T\Lambda_{\eta[m]k})^2\\
&\qquad +\kappa_s\sum_{k=1}^{q_1}\sum_{l}^{q_{(k)}}\left[\frac{\Lambda_{\eta kl}^2}{2|\Lambda_{\eta[m]kl}|}+\frac{|\Lambda_{\eta[m]kl}|}{2}\right] \;=\; g_s(\Lambda_\eta|\Lambda_{\eta[m]})
\end{aligned}
$$

Next, we should obtain the conditional expectation of the surrogate. This can be done using the following derivations.

$$
\begin{aligned}
&(\eta_{ki}-\frac{\omega_{(k)li}}{\alpha_{li}}(\Lambda_{\eta kl}-\Lambda_{\eta[m]kl})-\omega_{(k)i}^T\Lambda_{\eta[m]k})^2\\
&=\eta_{ki}^2+\frac{\omega_{(k)li}^2}{\alpha_{li}^2}(\Lambda_{\eta kl}-\Lambda_{\eta[m]kl})^2+\Lambda_{\eta[m]k}^T\omega_{(k)i}\omega_{(k)i}^T\Lambda_{\eta[m]k}-2\eta_{ki}\omega_{(k)i}^T\Lambda_{\eta[m]k}\\
&\qquad -2\eta_{ki}\frac{\omega_{(k)li}}{\alpha_{li}}(\Lambda_{\eta kl}-\Lambda_{\eta[m]kl})+2\frac{\omega_{(k)li}}{\alpha_{li}}(\Lambda_{\eta kl}-\Lambda_{\eta[m]kl})\omega_{(k)i}^T\Lambda_{\eta[m]k}
\end{aligned}
$$

Taking the expectation,

$$
\begin{aligned}
&E((\eta_{ki}-\frac{\omega_{(k)li}}{\alpha_{li}}(\Lambda_{\eta kl}-\Lambda_{\eta[m]kl})-\omega_{(k)i}^T\Lambda_{\eta[m]k})^2|Y,\theta)\\
&=E(\eta_{ki}^2|Y,\theta)+\frac{E(\omega_{(k)li}^2|Y,\theta)}{\alpha_{li}^2}(\Lambda_{\eta kl}-\Lambda_{\eta[m]kl})^2+\Lambda_{\eta[m]k}^TE(\omega_{(k)i}\omega_{(k)i}^T|Y,\theta)\Lambda_{\eta[m]k}\\
&\qquad -2E(\eta_{ki}\omega_{(k)i}^T|Y,\theta)\Lambda_{\eta[m]k}-2\frac{E(\eta_{ki}\omega_{(k)li}|Y,\theta)}{\alpha_{li}}(\Lambda_{\eta kl}-\Lambda_{\eta[m]kl})\\
&\qquad +2\frac{1}{\alpha_{li}}(\Lambda_{\eta kl}-\Lambda_{\eta[m]kl})E(\omega_{(k)li}\omega_{(k)i}^T|Y,\theta)\Lambda_{\eta[m]k}\\
&=W_{i|k,k}+\frac{[W_{i|\mathcal{A}_k}]_{l,l}}{\alpha_{li}^2}(\Lambda_{\eta kl}-\Lambda_{\eta[m]kl})^2+\Lambda_{\eta[m]k}^TW_{i|\mathcal{A}_k}\Lambda_{\eta[m]k}\\
&\qquad -2V_{\eta_k}^T\Lambda_{\eta[m]k}-2\frac{[V_{\eta_k}^T]_l}{\alpha_{li}}(\Lambda_{\eta kl}-\Lambda_{\eta[m]kl})+2\frac{1}{\alpha_{li}}(\Lambda_{\eta kl}-\Lambda_{\eta[m]kl})[W_{i|\mathcal{A}_k}]_l\Lambda_{\eta[m]k}
\end{aligned}
\qquad (4.2.37)
$$

Therefore,

$$
\begin{aligned}
&E(g_s(\Lambda_\eta|\Lambda_{\eta[m]})|Y,\theta) \\
&= \sum_{i=1}^{n}\sum_{j=1}^{q_1}\sum_{l}^{q_{(k)}} \frac{\alpha_{li}}{\psi_{\zeta k}}\Bigg[W_{i|k,k} + \frac{[W_{i|\mathcal{A}_k}]_{l,l}}{\alpha_{li}^2}(\Lambda_{\eta kl} - \Lambda_{\eta[m]kl})^2 + \Lambda_{\eta[m]k}^T W_{i|\mathcal{A}_k}\Lambda_{\eta[m]k} - 2V_{\eta_k}^T\Lambda_{\eta[m]k} \\
&\qquad\qquad - 2\frac{[V_{\eta_k}^T]_l}{\alpha_{li}}(\Lambda_{\eta kl} - \Lambda_{\eta[m]kl}) + 2\frac{1}{\alpha_{li}}(\Lambda_{\eta kl} - \Lambda_{\eta[m]kl})[W_{i|\mathcal{A}_k}]_l\Lambda_{\eta[m]k}\Bigg] \\
&\quad + \kappa_s\sum_{k=1}^{q_1}\sum_{l}^{q_{(k)}}\left[\frac{\Lambda_{\eta kl}^2}{2|\Lambda_{\eta[m]kl}|} + \frac{|\Lambda_{\eta[m]kl}|}{2}\right]
\end{aligned}
\tag{4.2.38}
$$

where $[W_{i|\mathcal{A}_k}]_{l,l}$ is (l,l) element of $W_{i|\mathcal{A}_k}$, $[V_{\eta_k}^T]_l$ is l-th element of $V_{\eta_k}^T$, and $[W_{i|\mathcal{A}_k}]_l$ is l-th column of $W_{i|\mathcal{A}_k}$.

Finally, differentiate the equation above, the solution can be achieved.

$$
\begin{aligned}
&\frac{\partial E(g_s(\Lambda_\eta|\Lambda_{\eta[m]})|Y,\theta)}{\partial\Lambda_{\eta kl}} \\
&= \sum_{i=1}^{n}\frac{\alpha_{li}}{\psi_{\zeta k}}\left[\frac{2[W_{i|\mathcal{A}_k}]_{l,l}}{\alpha_{li}^2}(\Lambda_{\eta kl} - \Lambda_{\eta[m]kl}) - \frac{2[V_{\eta_k}^T]_l}{\alpha_{li}} + \frac{2}{\alpha_{li}}[W_{i|\mathcal{A}_k}]_l\Lambda_{\eta[m]k}\right] + \kappa_s\frac{\Lambda_{\eta kl}}{|\Lambda_{\eta[m]kl}|} = 0 \\
&\Rightarrow \quad \Lambda_{\eta kl}\left[2\sum_{i=1}^{n}\frac{\alpha_{li}}{\psi_{\zeta k}}\frac{[W_{i|\mathcal{A}_k}]_{l,l}}{\alpha_{li}^2} + \frac{\kappa_s}{|\Lambda_{\eta[m]kl}|}\right] \\
&\qquad\qquad = \; 2\sum_{i=1}^{n}\frac{\alpha_{li}}{\psi_{\zeta k}}\left[\frac{[W_{i|\mathcal{A}_k}]_{l,l}}{\alpha_{li}^2}\Lambda_{\eta[m]kl} + \frac{[V_{\eta_k}^T]_l}{\alpha_{li}} - \frac{1}{\alpha_{li}}[W_{i|\mathcal{A}_k}]_l\Lambda_{\eta[m]k}\right]
\end{aligned}
$$

$$
\Rightarrow \; \hat{\Lambda}_{\eta[m+1]kl} \; = \; \hat{\Lambda}_{\eta kl} \; = \; \frac{\sum_{i=1}^{n}\left[\frac{[W_{i|\mathcal{A}_k}]_{l,l}}{\alpha_{li}}\Lambda_{\eta[m]kl} + [V_{\eta_k}^T]_l - [W_{i|\mathcal{A}_k}]_l\Lambda_{\eta[m]k}\right]}{\sum_{i=1}^{n}\frac{[W_{i|\mathcal{A}_k}]_{l,l}}{\alpha_{li}} + \frac{\kappa_s}{2|\Lambda_{\eta[m]kl}|}\psi_{\zeta k}} \tag{4.2.39}
$$

4.3 Further Issues in fitting Lasso SEM

Thus far, we develop an algorithm to implement L_1-regularization to SEM. However, that's not the whole thing we need to do for achieving our goal. There are a couple of additional issues that should be dealt with – Rescaling and standardization issues.

4.3.1 Rescaling Issue for the Measurement Model

As the first issue, we shall discuss about rescaling issue for the measurement model. This model, which is equal to the confirmatory factor analysis model, has several constraints on the matrix Λ. One of them is 'Identification Constraint', which set one path coefficient to 1 per each latent variable. This is on the purpose of 1) reducing the number of equations so that it becomes less than the available number of data elements. In SEM, we have $p^* = p(p+1)/2$ unique elements in a covariance matrix S. Thus the total number of parameters should not exceed it for the estimation to be carried out. Setting some coefficients to constant such as 1 can be an aid. 2) Also, the identification constraint has the practical meaning. Note that conceptually the factor doesn't have its own scale. Fixing one path coefficient to 1 per each factor is a solution to this problem. Using the scales of measurement variables corresponding to fixed coefficients, it becomes feasible to estimate factor variances and other coefficients relatively to these scales.

However, LARS and MM results do not include this kind of constraint. In other words, there is no guarantee that those methods produce an estimation result with fixing properly some coefficients to 1. The first purpose above may not be important, since LARS and MM are able to work without any reduction in the number of parameters to be estimated. But the second one is influential

in that the LARS and MM estimates may be unstably changing according to the scale of observed variables and estimated conditional covariance of latent variables.

Further discussion with respect to the standardization of SEM result will make this issue be understood more clearly. Note that there exist two type of parameter estimates in SEM – unstandardized and standardized. The unstandardized estimate is a direct output of SEM estimation, which includes constraints given in advance. Therefore unstandardized estimate of Λ contains 1 as its element per each column. Standardized estimate is a rescaled version of the previous one. By this procedure, we make the variance of latent variables be 1 and use the reciprocal of its original standard deviation to rescale all the other estimate. Rescaling is also carried out with a scale of the implied covariance matrix so that all the diagonal elements of $\hat{\Sigma}$ become 1 – that is, standardized estimates produce the implied 'correlation' matrix. The standardization formula for SEM is illustrated in Appendix B.

Now, consider the following question; what does the result of LARS and MM count as? unstandardized or standardized? At first, we eliminate the unstandardized estimate from our candidates since LARS and MM estimates do not take our constraints into account. Also they cannot be considered as standardized estimates in that the implied covariance matrix they reproduce does not has unit-diagonal elements. We may think about estimating $\hat{\Psi}_\epsilon$ by $\hat{\psi}_{\epsilon j} = 1 - diag(\hat{\Lambda}_L \hat{\Phi}_L \hat{\Lambda}_L^T)$, where the subindex L indicates that the estimates are obtained using LARS or MM algorithm. This always produces unit-diagonal implied covariance. However, it has no theoretical background and cannot be validated.

With strict regard for the above discussion, we are of opinion that the estimates attained from LARS and MM are not appropriate to be our final result.

However, it should be acknowledged that those methods can produce the result from optimizing the regularized conditional expectation of log-likelihood. And these estimates are proportional to effects from latent variables to their measurements. Therefore, even though the absolute values of LARS and MM estimates cannot be used as our final result, they can play the role as an intermediate result. Consequently, we suggest the following method to derive the final result of estimation using the LARS or MM estimates.

At first, Let $(\hat{\Lambda}_L, \hat{\Phi}_L)$ and $(\hat{\Lambda}_M, \hat{\Phi}_M)$ be the estimate of (Λ, Φ) attained by LARS and MM methods. For convenience, we denote only $(\hat{\Lambda}_L, \hat{\Phi}_L)$ henceforth. We shall derive $(\hat{\Lambda}, \hat{\Phi})$ such that i) $\hat{\Lambda}_L\hat{\Phi}_L\hat{\Lambda}_L^T = \hat{\Lambda}\hat{\Phi}\hat{\Lambda}^T$, ii) $\hat{\Lambda}$ contains the constraints given in advance.

Next, let $\hat{\Lambda}_{L,k}, k = 1, \ldots, q$ be the elements of LARS or MM estimate located in the position where the constraint are given in Λ beforehand. Note that the number of these elements are equal to that of factors, since each factor has one coefficient set to be 1. Using these things, define the scaling matrix $S_\Lambda = diag(1/\hat{\Lambda}_{L,k})$ for $\hat{\Lambda}_L$, and $S_\Phi = diag(\hat{\Lambda}_{L,k})$ for $\hat{\Phi}_L$. These are matrices of size $q \times q$.

Multiplying these matrices to their appropriate objects, we can obtain the result. First, compute $\hat{\Lambda} = \hat{\Lambda}_L S_\Lambda$. Then $\hat{\Lambda}$ has elements whose values are 1, at the position the constraint originally imposed in Λ. Secondly, $\hat{\Phi}$ can be computed as follows. This is quite complex. Here we use the temporary estimate of factor correlation matrix, $\hat{\mathrm{P}}^*_\omega$, which can be obtained by equation (4.2.20). As we described in the previous section, this is the ML estimates of factor correlation matrix computed using EM-algorithm, as proposed in Rubin and Thayer(1982). However, This cannot be generalized to SEM since covariances/correlations of endogenous latent variables are computed as functions of other parameters (1.1.10).

Pre- and post-multiplying $\hat{\mathrm{P}}^*_\omega$ by S_Φ derives the result, which we denote as $\hat{\Sigma}^R_\omega$. This is the factor covariance matrix properly rescaled by LARS estimates. Finally, extract the lower-right $q_2 \times q_2$ partition of $\hat{\Sigma}^R_\omega$. Since the location of the partition is corresponding to that of Φ in equation (1.1.10), this can be used as our final estimates of exogenous latent variable covariance matrix $\hat{\Phi}$. Note that the remaining parts of $\hat{\Sigma}^R_\omega$ cannot be appropriate estimates for corresponding parts of Σ_ω. This is because they do not satisfy the relationships with other parameters contained in equation (1.1.10). Therefore, estimates for these parts should be obtained by summing and multiplying the estimates of B, Γ and Ψ_ζ.

The following box summarizes the above computation.

Step 1: Compute $(\hat{\Lambda}_L, \hat{\Phi}_L)$ or $(\hat{\Lambda}_M, \hat{\Phi}_M)$ using LARS or MM.
WLOG, we illustrate the following using LARS estimates.

Step 2: Calculate the rescaling matrices as follows.
$S_\Lambda = diag(1/\hat{\Lambda}_{L,k}), \quad S_\Phi = diag(\hat{\Lambda}_{L,k})$

Step 3: Obtain $\hat{\Lambda} = \hat{\Lambda}_L S_\Lambda$.

Step 4: i) $\hat{\mathrm{P}}^*_\omega \leftarrow$ Standardize $W^* = [W + W^T]/2$
ii) Compute $\hat{\Sigma}^R_\omega = S_\Phi \hat{\mathrm{P}}^*_\omega S_\Phi$
iii) Extract the lower-right $q_2 \times q_2$ partition of $\hat{\Sigma}^R_\omega$.
This partition can be used as $\hat{\Phi}$.

$\Rightarrow$ The result estimates are $(\hat{\Lambda}, \hat{\Phi})$.

4.3.2 A Standardization Issue in M-step

In the previous subsection, our discussion covers the standardization of SEM estimation results. However, standardization of variables is also an important issue. Here, standardization of variables indicates that we center and scale our variables before we use them in our main analysis. It should be noted that standardization of results and that of variables are completely different in SEM, while they are conceptually the same in linear regression analysis. In fact, standardized variables bring about standardized result in regression models, but this is not true for SEM. In this subsection, we shall discuss the standardization of variables in SEM regarding its necessity and the method to carry it out in the LARS and MM-algorithm.

When a linear regression model is analyzed using Lasso, one of the most important issue is the standardization of variables. Since Lasso regularizes a loss function with the sum of absolute values of coefficients, its estimates are influenced by scales of regressor variables. Therefore, if researchers hope to remove these scale effects and investigate only the pure contributions of each variable to Lasso estimates, standardization should be conducted before fitting Lasso. When he first suggested Lasso, Tibshirani(1996) stated that all the regressor variables were standardized and the response variable was centered. Most of his follow-up researches have conformed this statement, except only when scales of the regressor variables have significant meaning. In fact, centering is not required since means of regressors do not affect the sizes of coefficients. Therefore, sometimes we shall use the term 'standardization' in the meaning of scaling.

In Structural Equation Modeling, however, this issue becomes quite an abstruse problem. Note that in this model regressors in both of measurement and

structural models are latent, unobserved variables. Then, is it appropriate to standardize latent variables whose scale is not defined? Furthermore how can we standardize these variables?

With regard to the first question, consider that it is the conditional expectation and covariance of latent variables that we use to estimate SEM by LARS or MM-algorithm. If we want to standardize the regressor variables as in the linear regression, we should deal with those two terms. But what we have to bear in mind is that they are also estimated during the estimation procedure. This means that their scales are naturally computed in the process of SEM estimation, thus manipulating them may cause poor investigation into the deep structure among the variables.

The Measurement model may not be the problem since it deals with observed variables as its response and latent variables as its regressors. Though observed variables are involved in the calculation of conditional moments of latent variables, there is no valid reason not to carry out the standardization. Also, since the LARS and MM estimates are used as intermediate results in rescaling the path coefficient, it is important to extract pure relationships between observed responses and factors by standardization.

However, in the structural model our responses and regressors, all of which are latent, have inseparable relations among them. Note that each endogenous latent variable is a response variable for each corresponding equation in the structural model. However this variable can also be one of independent variables in another equation. Additionally, exogenous latent variables act as regressors for each equations in the structural model, sometimes for more than one equations. The point is that latent variables are intricately entwined with each other in SEM. Therefore it is difficult to think each of latent variables independently with the other factors.

Furthermore, it should be noted that the conditional moment vector and matrix of these latent variables are estimated together during SEM estimation. Thus these results reflect the naturally extracted dynamics among these factors. If we try to standardize variables in the structural model, we should carry it out to each of q_1-equations with changing a response and regressors adequately to the equation analyzed at the moment. Therefore, each latent variable is standardized in some equations and is not in other equations. Also in each equation in the structural model, only regressors are standardized and the corresponding response remains raw, even though scales of all of these latent variables are estimated together so that those scales are appropriately determined. In this regard, standardization for the structural model seems not to produce any positive effect, only to distort underlying relation.

This conclusion is a result from our subjective inference. Thus this should be supported by careful scrutiny regarding the mathematical formula of SEM and its process of estimation. Nonetheless, our small pre-simulation analysis reveals that among the four possible cases regarding the combination of standardization/unstandardization of measurement/structural models, our method yields the most stable and plausible result. Hence we shall follow this conclusion in the main simulation analysis in the present thesis, bearing in mind that this needs further validation.

For the next step, we shall deal with the second question – the method to standardize regressor variables. The main idea is centering the conditional expectation and scaling the conditional variances of latent regressor variables. For the purpose of fitting the SEM, LARS programs require the result terms in (4.2.25) to solve the problem (4.2.19). These terms, $\tilde{\eta}$ and $\tilde{\Omega}$, contain $W_{\mathcal{A}_k}$ and V_{η_k}, which are the partitions of the matrix W. Note that W contains conditional expectation of 2nd moments and joint moments among all the latent

variables in the model. When these variables are considered to be standardized, W can be regarded as their conditional variance estimates. Therefore, the idea is rescaling this matrix so that its diagonal elements are to be 1, and off-diagonal elements are to be adjusted properly. After this procedure, the partitions corresponding to $W_{\mathcal{A}_k}$ and V_{η_k} are to be extracted and defined as $W^S_{\mathcal{A}_k}$ and $V^S_{\eta_k}$, which denotes standardized $W_{\mathcal{A}_k}$ and V_{η_k}.

The following box explains the entire procedure, for the case when a LARS program is used to estimate the structural model in Lasso SEM.

Standardization in Lasso SEM

Case : Structural Model using LARS

Step 1: Rescale the matrix W.

$$W^S \stackrel{\text{def}}{\equiv} diag(W)^{-\frac{1}{2}}\, W\, diag(W)^{-\frac{1}{2}}$$

Step 2: Extract related matrices.

$$W^S_{\mathcal{A}_k} \leftarrow I_{q(\mathcal{A}_k^c)} W^S I^T_{q(\mathcal{A}_k^c)}$$
$$(V^S_{\eta_k})^T \leftarrow I_{q(\{k\}^c)} W^S I^T_{q(\mathcal{A}_k^c)}$$

Step 3: Input these matrices to equation (4.2.19).

Other cases – when MM-algorithm is exploited, or when we also carry out standardization for the measurement model, are the simple modifications of the above.

4.3.3 Tuning Methods for L_1-Regularized SEM

In Lasso, or other regularization methods, the value of tuning parameter κ should be determined before estimation. When dealing with the linear regression models, we can make use of several cross-validation methods such as K-fold Cross Validation, LOOCV(Leave-One-Out Cross Validation), and so on(Hastie, Tibshirani & Freedman, 2008). Those procedures separate sample data into several training sets and a test set. Training sets are used to obtain estimates with the given value of κ, and the estimated prediction error (4.3.1) is calculated for the test set.

$$\widehat{Err}_k = \sum_i (y_i - \hat{y}_i)^2 \tag{4.3.1}$$

When K-fold method is used, sample is divided into K equal parts and one of these is used as the test set. After the prediction error for this set is calculated, other set is selected as the next test set and the others, including the former test set, are regarded as the training sets. Therefore, K prediction errors are obtained and averaged to be used as the criterion to determine whether the input κ is good enough or not. This criterion is called the cross-validation estimate of prediction error.

$$CV(y, \hat{y}) = \frac{1}{K} \sum_{k=1}^{K} \widehat{Err}_k \tag{4.3.2}$$

LOOCV can be understood as the modification of K-fold CV with $K = n$. And among the candidates of κ, the value minimizing the cross-validation estimate of prediction error is determined to the optimal tuning parameter.

However, this methods can not be applied to the structural equation modeling. It is strongly emphasized that SEM is the model for explanation on real-world phenomenon, not for the prediction on future data. In fact, since the

regressor variables are latent and unobserved in SEM, it is not be able to be used to predict future measurement variables. Hence it is impossible to calculate the criterion (4.3.2) for SEM.

Despite this problem, the fundamental logic of the cross-validation is still applicable to SEM as long as another suitable criterion is given. One of the most promising candidate indices is OD. As described in the earlier chapters, this value offers the degree of discrepancy between the population covariance/correlation matrix and the implied covariance/correlation matrix. Low values of OD indicates that the current model is excellent in explaining the interested phenomenon or covariance structure, and is also reproducible or generalizable across independent sample sets.

One big problem this suggestion encounters is that OD can be calculated only when we know the population covariance/correlation matrix. This is practically impossible in most of researches. Alternatively we are going to suggest a new index, which acts as an estimate of OD and can be calculated only with sample data. The idea originates from the logic of cross-validation, with a little variation. Assume that A group of researchers has the hypothesized model estimated using their own sample data. However they are concerned about a generalizability of their model. To study further on this issue, they obtain K distinct data sets which are sampled from the same population with the original one. Then it is able to obtain the implied covariance matrix $\hat{\Sigma}$ from the original fitting result, and K sample covariance matrix, $S_1, S_2, \ldots, S_K$ from the new sample sets. Define ASD_k or 'Across-sample Discrepancy with respect to k-th sample' as follows.

$$ASD_k \stackrel{\text{def}}{\equiv} F(S_k, \hat{\Sigma}), \quad k = 1, 2, \ldots, K. \tag{4.3.3}$$

Note that this is a generalization of cross validation index(Cudeck & Browne,

1983; Browne & Cudeck, 1989). Type of the discrepancy function should be determined in accordance with researchers' purposes of analysis. Using these K-ASD_k's, we can compute the following index which indicates an average discrepancy between the original training sample and each of the other samples obtained afterward. This can be used as an estimate of overall discrepancy. We denote this index as 'ASD(Across-sample Discrepancy)' or 'SD.cv(Cross-validated Sample Discrepancy)'.

$$ASD \stackrel{\text{def}}{\equiv} \frac{1}{K} \sum_{i=1}^{K} ASD_k \tag{4.3.4}$$

Readers may doubt the presumption stating that K sample sets distinct from the original training set are prepared in advance since it is difficult for researchers to make provision for this large scale data. In order to relieve this, some other strategies can be applied to our practical situation. For instance, the original K-fold method can be applied directly to our data set. Or K-new sample sets for generalization can be obtained by resampling methods such as Bootstraping.

4.4 Result Algorithm for Lasso SEM

The Lasso SEM, L_1-regularized Structural Equation Modeling, described in this chapter can be abbreviated as the following algorithm flow.

Step 1.
In the $[m+1]$-th step, let $\hat{\theta}_{[m]}$ be the vector of estimates obtained from the previous step. Compute the conditional distribution of ω given Y and $\hat{\theta}_{[m]}$, as in (4.2.4). This yields δ and Δ.

Step 2.
Compute W and $W_{(j)}$ in (4.2.6) and $W_{k,k}, V_{\eta k}$, and $W_{\mathcal{A}k}$ in (4.2.10) $\sim$ (4.2.12).

Step 3.
Using the results in step 1 and 2, derive $RELL_M$ and $RELL_S$ in (4.2.14) and (4.2.15).

Step 4.
Compute $\hat{\Psi}_\epsilon$ and $\hat{\Lambda}_L$ by minimizing $-2RELL_M$. For $\hat{\Psi}_\epsilon$, refer to the formula (4.2.16). And for $\hat{\Lambda}_L$, LARS or MM can be exploited as in Section 4.2.3. Standardization procedure can be added as described in Section 4.3.2., if it is deemed necessary.

Step 5.
Obtain $\hat{\Lambda}$ by rescaling described in Section 4.3.1.

Step 6.
Compute $\hat{\mathrm{P}}^*_\omega$ in (4.2.20) and rescale it as described in Section 4.3.1. This step yields $\hat{\Phi}$.

Step 7.

Compute $\hat{\Psi}_\zeta$ and $\hat{\Lambda}_\omega$ by minimizing $-2RELL_S$. For $\hat{\Psi}_\zeta$, refer to the formula (4.2.18). And for $\hat{\Lambda}_\omega$, LARS or MM can be exploited as in Section 4.2.3. Standardization procedure can be added as described in Section 4.3.2., if it is deemed necessary.

Step 8.

Repeat step 1-7 until the convergence of estimates is achieved. As in Choi(2010), the convergence can be considered only with respect to the factor loadings and regression coefficients.

After the iteration is over,

Step a.

$\hat{\Sigma}_\omega$, the estimate of covariance matrix of the latent variables can be computed using the formula (1.1.10).

Step b.

Also the implied covariance matrix can be computed as $\hat{\Sigma} = \hat{\Lambda}\hat{\Sigma}_\omega\hat{\Lambda} + \hat{\Psi}_\epsilon$.

Lasso for the factor analysis model is the special case of the above algorithm, considering only the $RELL_M$ and corresponding parameters. In Appendix D, an R function for Lasso SEM is presented.

Chapter 5

Simulation Study : Method

5.1 Purposes of Research

The present thesis conducts several simulation researches to study effects of Lasso application to the structural equation modeling. More specifically, our main points are abbreviated in the following questions.

1) *Do the Lasso's capability of shrinkage estimation and variable deletion also appear in SEM? That is, is it possible to obtain more parsimonious and sparse SEM results by Lasso regularization?*
2) *Can the Lasso reduce the generalizability/reproducibility indices such as OD or MSE when it is applied to SEM?*

Additionally, we shall find answers for the following questions in order to obtain more meaningful understanding on regularization and generalizability in the structural equation modeling.

3) *Do BLasso estimates not shrink completely to zero? How about considering them to be zero if their values are less than a pre-determined bound?*
4) *Do overall discrepancy and mean squared error have close relationships empirically?*

5) *Is the existing optimization method of SEM, which is based on minimizing the sample discrepancy, able to produce a generalizable outcome? That is, does the sample discrepancy have high correlation with the overall discrepancy?*

6) *Among the model fit indices generally used in SEM, which of them can play a role as an indicator of generalizability and reproducibility? That is, which of model fit indices show empirically meaningful relationships with OD or MSE?*

In order to find solutions to these questions, the comparison between Lasso SEM and MLE, the most dominant estimating method for SEM, is to be conducted by the Monte Carlo simulation. BLasso is also fitted and investigated with these two methods for several purposes. The most important criteria of this comparison analysis are the indices related to generalizability and reproducibility, such as OD and MSE. And also some indices including discrepancy statistics, variance, bias squared, and several model fit indices are computed too. Since some of these values requires the population analysis results in their computation, the simulation study should start with generating population. And by simple random sampling from this population, we can obtain sample data sets to be used for analyses in our simulation.

It should be considered that there is a possibility that Lasso and ML perform differently according to several conditions, such as types of model analyzed, sample size, and values of the covariance matrix. Therefore, in order to take various situations into account, the analysis proceeded with manipulating those conditions.

Among the available algorithms for Lasso SEM optimization, the LARS algorithm is selected for present simulation studies. This choice is for the practical reasons such as simulation time and analytical convenience, but in our small pre-simulation study MM-algorithm also performs well.

5.2 Generating Population

Before illustrating the models and conditions in our simulation, we shall describe how we generate population data for each of research designs. The population covariance matrix can be conceptualized in various ways. Out of all the candidates, those three illustrated in Section 1.4 are most well-organized concepts(Cudeck & Henly, 1991; Cudeck & Browne, 1992).

a) $\Sigma_0 =$ arbitrary covariance matrix

b) $\Sigma_0 = \Sigma_0(\theta_0)$

c) $\Sigma_0^* = \Sigma_0(\theta_0) + E$

Among these concepts, researchers who attempts to conduct a simulation study on SEM can exploit b). It's easy to apply this concept - what we have to do is just specifying the parameter matrices, and compute the covariance matrix using the formula (1.1.10) and (1.2.1). This notion has been widely used in studying SEM(For example, Hu & Bentler, 1998). As we mentioned in the earlier chapter, in the notion of b), there exists an operating model $\Sigma_0(\cdot)$ that can produce Σ_0 exactly. However, it is implausible to say that this model can be discovered in practice, since the model contains too much complexity. Nevertheless, it is worth studying this concept by Monte Carlo simulation, under the assumption that the operating model we used in generating the population is unidentifiable.

From a different standpoint, c) has been accepted as theoretically the most valid and reasonable concept(Tucker, Koopman & Linn, 1969; Cudeck & Henly, 1991; Cudeck & Browne, 1992; Cheung & Rensvold, 2001; Cheung & Rensvold, 2002; Coffman, 2008; Chun & Shapiro, 2010; MacCallum, Lee & Browne, 2010). This argument can be supported considering the mechanism implied in real

world data generating process. The population can be assumed to be affected by the three types of factors; major factors, unique factors, and minor factors. The major factor indicates one that gives meaningful and significant contribution on our population. These variables connected strongly with the true-model. And the relationship between major factors and data-generating process are usually simple and direct. Since they can account for a huge proportion of variation in the population, mostly main purposes of scientific researches lie in finding and comprehending the system of major factors related to the interested phenomenon. In SEM, we can deal with these factors as latent variables or by manifest variables. Also we can model the system of structural equations after our hypothetic interaction and connection among those variables.

Unique factors can also be included our model. These factors contain various types of error only related to each variable. Error terms in SEM models are reflections of this kind of factor. Also note that 'uniqueness' in factor analysis model refers to the variance assumed to be explained by these unique factors in proportion to the total variance of respective measurement variables. Measurement error is a notable example of unique factors. And many other randomness in the data can be designed as unique factors. In most cases, these factors are assumed to be independent with each other.

The last type of factor is the minor factor that refers to all the factors potentially affecting the true data-generating process. Previous researches(Cudeck & Henly, 1991; Cudeck & Browne, 1992) suggest non-random error and localized method effects as examples for this type of factor. However, we are of opinion that more broader effects can be categorized in to minor factors if they have only paltry and meaningless effects on the population. Note that there is a possibility that any two variables are related to each other even though

seemingly they are not. In spite that we cannot find any connection among some variables, maybe it happens in the deep underlying process but to a negligible degree. Those minor effects can also be modeled as one of variables in SEM, as long as researchers are able to find any plausible theory or hypothesis. And this may be able to reduce the model error contained in matrix E and discrepancy due to approximation. However, an increase in model complexity ensue as one of the consequences. One paramount difference between the major factor and the minor factor may lie in this point in that minor factors cannot contribute enough to the model to such an extent to offset the inflated model complexity, while major factors can do. In this regard, to consider these factors into account only brings about some problems similar with overfitting, poor generalizability and interpretability. Therefore, this can cause a hindrance to the quasi-true model $\Sigma_0(\cdot)$ in approximating the corresponding data-generating process. Hence, the minor factors are supposed not to be included in the best approximating model.

Now, returning to the three conceptualization of the population generating process, the concept c) is most pertinent to embrace those three types of factors and their dynamics. The major and unique factors can be modeled as the best approximating model, but they cannot reproduce the true data-generating mechanism perfectly. The difference which results from a huge number of minor factors can be represented in matrix E. Many of previous researches discuss the importance of this conceptualization in Monte Carlo experiment studies(MacCallum et al, 1994; Olsson, Foss & Breivik, 2004; Tomarken & Waller, 2005), and some studies are carried out following this conceptualization to obtain valid and reliable results from their simulation studies(Cudeck & Browne, 1992; Cheung & Rensvold, 2001; Cheung & Rensvold, 2002; Coffman, 2008; Chun & Shapiro, 2010; MacCallum, Lee & Browne, 2010).

The key remaining task is to find the method to generate population reflecting the conceptualization c). The most important problem is that we should be able to manipulate the degree of perturbation so that the effect of matrix E is in an appropriate degree to be regarded as a result from minor factors. Also, this matrix E should not be involved in determination of values of θ_0. This condition enables the minimization of discrepancy between Σ_0^* and $\Sigma_0(\theta_0)$ to be achieved by true values of θ_0.

In fact, Cudeck and Browne(1992) proposed the excellent procedure for implementing the task. Their method derives the matrix E and adds it to $\Sigma_0(\theta_0)$, producing Σ_0^*. The degree of model error can be controlled by δ, the scalar-valued parameter used in generating E. δ-insertion can be done based on various types of discrepancy function including OLS, GLS, and ML. Estimation using the resulting covariance matrix is supposed to recover θ_0 and δ exactly when the model is specified correctly as the quasi-true model and the estimating method exploits the same discrepancy function used in the derivation of model error matrix. As they introduce their method gently in detail, only the summary of the procedure will be presented in Appendix E.

In the present thesis, we shall carry out the simulation studies with both b) and c) in conceptualizing the population covariance matrix. This effort will give us the insight on the effect of model error on the performance of estimation methods investigated in the studies.

5.3 Research Models

Simulations are conducted for the confirmatory factor analysis model and the structural equation model separately. Usually CFA model is regarded as a special case of SEM. Since the former contains essential features of the latter model, results of CFA simulation study are understood as it can be generalized to SEM models in many cases.

However, there are critical differences between these models. For example, CFA model contains only the latent exogenous variables. Therefore in this model, latent variables are specified purely as regressor variables. In SEM, on the other hand, there exist the latent endogenous variables, which are affected by and also affect other latent variables as noted in Chapter 1. Hence, they act as regressors in some equations and as responses in other equations, making intricate interactions among those latent variables. This structural difference generates fundamental distinction between CFA and SEM.

Also, due to the consideration on the structural model which contains hypothetic relationships among latent variables, SEM contains far more parameters than the CFA model. In other words, SEM has fairly higher model complexity than CFA. Note that in general, SEM includes many latent variables that each of them affect a lot of corresponding measurement variables. Furthermore these measurement variables and latent endogenous variables in the structural models are also affected by their own unique factors. This complicated structure in SEM often should be considered as disparate to that of CFA model.

Hence, all things considered, we shall study CFA and SEM independently in the next chapter.

In some parts of parameter specifications, we referred to Hu and Bentler's

method. Studying the CFA model, they set variances of measurement variables to be 1. To be specific, with given Λ and Φ, unique variances in the matrix Ψ_ϵ is calculated as follows so that they would yield unit-variances for measurement variables(Hu Bentler, 1998).

$$\Psi_\epsilon = I_p - diag(\Lambda\Phi\Lambda^T) \tag{5.3.1}$$

When it comes to SEM models, Ψ_ϵ can be determined in the same way with replacing Φ by Σ_ω. Note that Σ_ω can be computed using parameters values in the model, by equation (1.1.10).

For each type of models, several individual models and its parameter values are specified as follows.

1) Factor Analysis Model

At first, the factor loading matrix Λ is given as follows.

$$\mathbf{\Lambda} = \begin{pmatrix} 1 & 0.92 & 0.90 & 1.05 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0.98 & 1.10 & 0.90 \end{pmatrix}^T$$

$$\Phi = \Phi_m, \quad \text{where} \quad m = 1, \ldots, 6.$$

$$\Psi_\epsilon = I_p - \Lambda\Phi\Lambda^T$$

In this specification, we include the identification constraint. This constraint is necessary in CFA and SEM estimation to make the convergence of iteration be possible. As illustrated in the previous chapter, this constraint conceptually aims to set the scales of latent variables which originally have no scales, using the observed scale of the one selected measurement variable per each latent. Usually, this is done by setting the path coefficients from each latent variable to their first measurement variable to be 1 as identification constraints. For example, in the above specification of Λ, $\lambda_{11} = \lambda_{51} = 1$.

It should be noted that even though all values in Λ is fixed and unchanged during all the simulation studies, this matrix can produce sample covariance matrices with elements of different sizes, in combination with various Φ matrices. Values in Φ matrix will be discussed below as one of research conditions.

The following figure represents individual CFA models used in our research.

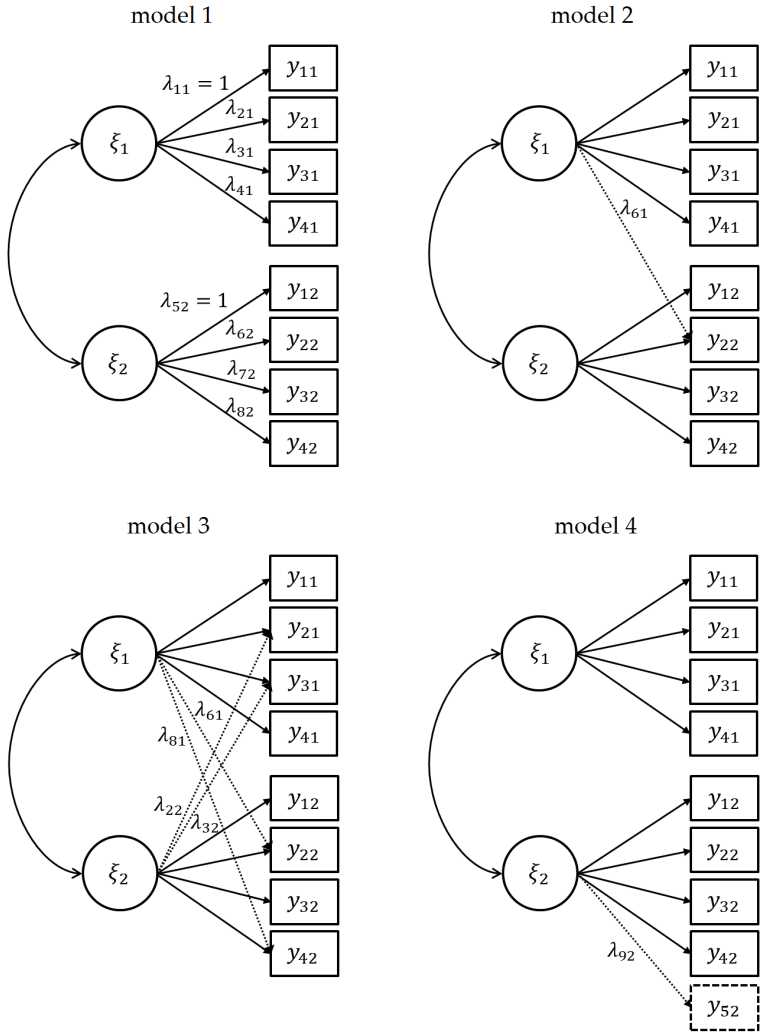

Figure 5.1: Factor Analysis Models in the Present Simulation Study

These models are chosen to reflect a variety of situations that we may encounter when analyzing CFA models. The list in the below abbreviates respective meanings of the models.

Model 1) The model is correctly specified.
Model 2) The model is slightly misspecified.
Model 3) The model is moderately misspecified.
Model 4) The model includes an irrelevant variable.

The model 1 is the one we will use in generating population. That is, model 1 represent $\Sigma_0(\cdot)$. Thus, this implies the case when we correctly specified the best approximating model. In reality it's not supposed to happen. Also note that even if we know $\Sigma_0(\cdot)$ and parameter values of θ_0 completely, what we can achieve is $\Sigma_0(\theta_0)$, which is only an approximation of Σ_0. This model is included in our analysis to obtain the baseline result; No misspecification so that no complete shrinkage occurs even though the Lasso in implied.

The other three models represent the case we misspecifies the mode. Model 2 and 3 have additional misspecified paths compared to model 1, but differ in the degree of misspecification. Lasso is expected to remove those paths so that leads us more closely to $\Sigma_0(\cdot)$. Furthermore, it is also anticipated that Lasso performs relatively better in model 3 than model 2 in comparison with the ML method, since the former model has more specification.

We expect the same effect for model 4. However this model has a slightly different type of error in model specification. While the former two models have additional paths and maintain their matrix size, model 4 has additional variables, which is irrelevant to the generating process. This case represents that some minor variables are included wrongly or accidentally in the model.

Hence, a result from Lasso may contain different features.

The above models can be described in mathematical forms as follows.

For j = 1, . . . , p, k = 1, . . . , q (Here, p = 8, q = 2)

Model 1) $y_j = \lambda_{j1}\omega_1 + \epsilon_j, \quad \text{for} \quad j = 1, \ldots, 4$

$y_j = \lambda_{j2}\omega_2 + \epsilon_j, \quad \text{for} \quad j = 5, \ldots, 8$

Model 2) Same as model 1, except

$y_j = \lambda_{j1}\omega_1 + \lambda_{j2}\omega_2 + \epsilon_j, \quad \text{for} \quad j = 6$

Model 3) Same as model 1, except

$y_j = \lambda_{j1}\omega_1 + \lambda_{j2}\omega_2 + \epsilon_j, \quad \text{for} \quad j = 2, 3, 6, 8$

Model 4) Same as model 1, but add

$y_9 = \lambda_{92}\omega_2 + \epsilon_9$

2) Structural Equation Modeling

Analogously to FA model, parameter values in coefficient matrices, Λ_y, Λ_x, B, and Γ are given as follows. A perturbation matrix for latent endogenous variables, Ψ_ζ, is also presented.

$$\mathbf{\Lambda_y} = \begin{pmatrix} 1 & 0.92 & 0.90 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0.85 & 0.93 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0.93 & 1.02 \end{pmatrix}^T$$

$$\mathbf{\Lambda_x} = \begin{pmatrix} 1 & 0.89 & 1.05 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0.97 & 0.99 \end{pmatrix}^T$$

$$\mathbf{B} = \begin{pmatrix} 0 & 0 & 0 \\ 0.45 & 0 & 0 \\ 0 & 0.65 & 0 \end{pmatrix} \qquad \mathbf{\Gamma} = \begin{pmatrix} 0.70 & 0 \\ 0 & 0.55 \\ 0 & 0 \end{pmatrix}$$

$$\mathbf{\Psi}_\zeta = \begin{pmatrix} 0.3 & 0 & 0 \\ 0 & 0.25 & 0 \\ 0 & 0 & 0.28 \end{pmatrix}$$

Note that B is lower triangular and Ψ_ζ is diagonal. This reflects that the best approximating model $\Sigma_0(\cdot)$ satisfies recursiveness.

The following figure shows SEM models we will deal with in simulation researches. These models represent various cases as in CFA part. Notice that the manipulation is carried out to the matrix Γ.

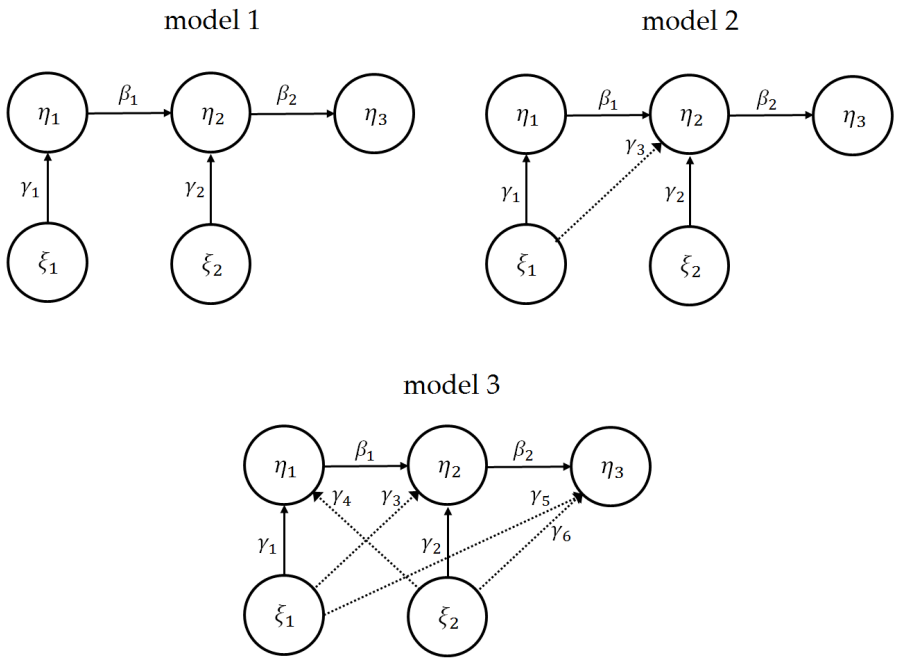

Figure 5.2: Structural Equation Models in the Present Simulation Study

Model 1) The model is correctly specified.

Model 2) The model is slightly misspecified.

Model 3) The model is moderately misspecified.

Model 1, 2, and 3 is chosen in the same way as previous CFA case. Model 1 represents the quasi-true generating model, and the other two has additional misspecification to varying degrees. In SEM, misspecification can occur in various ways; for $\Lambda_y, \Lambda_x, B,$ and Γ. Also, more complex model may contains additional covariances between any two variable in the model, including unique factors. However, in the present thesis, we shall only deal with the case that misspecification is added to matrix Γ for simplicity. When there exist additional misspecification in the measurement model or other equations in the structural model, it is expected that Lasso performs better with shrinking those coefficients completely to zero. After demonstrating Lasso's performance in this case, more general cases should be investigated.

Those models can be expressed in mathematical equations as follows.

Measurement Model

$$
\begin{aligned}
y_j &= \lambda_{y_{j1}}\eta_1 + \epsilon_j, \quad \text{for} \quad j = 1,\ 2,\ 3 \\
y_j &= \lambda_{y_{j2}}\eta_2 + \epsilon_j, \quad \text{for} \quad j = 4,\ 5,\ 6 \\
y_j &= \lambda_{y_{j3}}\eta_3 + \epsilon_j, \quad \text{for} \quad j = 7,\ 8,\ 9 \\
x_j &= \lambda_{x_{j1}}\xi_1 + \epsilon_j, \quad \text{for} \quad j = 1,\ 2,\ 3 \\
x_j &= \lambda_{x_{j2}}\xi_2 + \epsilon_j, \quad \text{for} \quad j = 4,\ 5,\ 6
\end{aligned}
$$

Since we don't consider any misspecification in the measurement model, the above equations are common in all the models involved. No misspecifica-

tion in the measurement model may seem implausible in practice. However, sometimes this can be regarded as true when the measurement variables are chosen from proper questionnaires validated in advance.

Structural Model

$$
\begin{aligned}
\text{Model 1)} \quad & \eta_1 = \gamma_1 \xi_1 + \zeta_1 \\
& \eta_2 = \beta_1 \eta_1 + \gamma_2 \xi_2 + \zeta_2 \\
& \eta_3 = \beta_2 \eta_2 + \zeta_3 \\
\text{Model 2)} \quad & \eta_1 = \gamma_1 \xi_1 + \zeta_1 \\
& \eta_2 = \beta_1 \eta_1 + \gamma_3 \xi_1 + \gamma_2 \xi_2 + \zeta_2 \\
& \eta_3 = \beta_2 \eta_2 + \zeta_3 \\
\text{Model 3)} \quad & \eta_1 = \gamma_1 \xi_1 + \gamma_4 \xi_2 + \zeta_1 \\
& \eta_2 = \beta_1 \eta_1 + \gamma_3 \xi_1 + \gamma_2 \xi_2 + \zeta_2 \\
& \eta_3 = \beta_2 \eta_2 + \gamma_5 \xi_1 + \gamma_6 \xi_2 + \zeta_3
\end{aligned}
$$

As described before, misspecifications of the structural model occurs only in the matrix Γ. Among those parameters, only γ_1 and γ_2 have nonzero true values. And the remaining, γ_3, γ_4, γ_5 and γ_6 are zero in their true values.

$$
\mathbf{\Gamma} = \begin{pmatrix} \gamma_1 & 0 \\ 0 & \gamma_2 \\ 0 & 0 \end{pmatrix} \quad \text{for model 1,} \qquad \mathbf{\Gamma} = \begin{pmatrix} \gamma_1 & 0 \\ \gamma_3 & \gamma_2 \\ 0 & 0 \end{pmatrix} \quad \text{for model 2,}
$$

$$
\text{and} \quad \mathbf{\Gamma} = \begin{pmatrix} \gamma_1 & \gamma_4 \\ \gamma_3 & \gamma_2 \\ \gamma_5 & \gamma_6 \end{pmatrix} \quad \text{for model 3.}
$$

5.4 Research Conditions

1) Sample Size

MLE's performance may differ depending on the size of the sample used in estimation since a lot of its excellent properties are based on asymptotic theories. Moreover, it is well known that results and estimates of Bayesian analyses, including BLasso SEM, converge to MLE as the sample size is getting larger. This is due to the fact that posterior distribution is a combination of prior distribution and likelihood function. Thus, if the sample size is large enough, contribution of the likelihood function could overwhelm that of prior in computing their posterior distribution. In addition, Lasso SEM algorithm presented in this thesis may also perform differently as sample size is varied.

Considering this point, different sizes of simple random samples are produced from the pre-generated population. Sizes are determined pursuant to those are being used or can be obtained by researchers inn ordinary studies. For CFA, whose model complexity is relatively low, conditions on the sample size are determined 50, 100, 200, and 1000. 50 may seem a little small, but this stand for the case that iterations of estimating methods are barely be able to converge. 100 and 200 are the commonly reported sample sizes in psychological researches using factor analysis model. And 1000 is quite a large size so that only a few researchers gained and used in their studies.

Sample sizes for SEM are determined similarly. However, in consideration of its higher model complexity compared to CFA model, a little larger sizes of samples are allocated – 150, 250, 500, and 1000. 150 is the size required for MLE estimation to converge in our research models. Size 100 was also a candidate, but this turned to be so small that more than half of iterations failed to converge. The other sizes are determined analogously to the CFA model case.

2) Covariance Matrix of Latent Variables

Data, or its covariance matrix contains the covariance/correlation structure among measurement variables, which is influenced by several latent variables. When the covariance/correlation structure comes out evidently into the data or its covariance matrix, the result of CFA or SEM estimation is usually nice. When this is not the case, a bad result also ensues. In addition some previous researches on SEM reveals the case when the CFA or SEM fits to the sample data quite well or not. For example, Browne et al.(2002) analyzed a covariance matrix with a considerably evident structure. However, even though all the elements in the residual matrix are very small, most of ML-based fit indices were poor. That is, in some cases, excessively high correlations among the variables might be able to produce a poor ML result.

The point is that the performance of estimation methods for CFA and SEM are affected by the degree of covariance/correlation among the interested measurement variable. Therefore, several data sets with a variety of degree of covariance/correlation should be tested in order to evaluate the performances of MLE, Lasso, and BLasso.

Since sample data and sample covariance matrix are obtained from their population, we are able to deal with several conditions for covariance/correlation structure by manipulating the population covariance/correlation matrix. However, in CFA or SEM, this covariance/correlation structure is assumed to be a matrix-valued function of lots of parameters such as $\Lambda, \Phi, \Psi_\epsilon$, etc. Therefore, if we manipulate the population covariance/correlation matrix directly, we lose control of these individual parameters. Hence it is recommended to take proper conditions on individual parameters, not on their output covariance matrix, into account.

As we described earlier, Ψ_ϵ is determined by the equation (5.3.1). And even

if we fix the parameter values of Λ, Λ_η and Ψ_ζ, various Σ_ω matrices and the population covariance/correlation matrices can be obtained simply by manipulating the Φ matrix. Therefore, in simulations of the present thesis, we analyze samples from different population covariance/correlation matrices derived from different Φ matrices and the other fixed parameter matrices.

The conditions for Φ matrix are as follows. Since we make the measurement variables have unit-variances and include the identification constraints in specifying Λ, factor variances cannot take values larger than 1. Otherwise, negative unique variance problems occur for the elements of Ψ_ϵ corresponding to those of Λ which have identification constraints. Thus $0 \leq \phi_{k,k} \leq 1$ for all k. Within this domain, 0.8, 0.6 and 0.4 are selected, to stand for the cases when the factor variances are large, moderate, and small. In addition we also deal with the cases when the covariances among latent variables are large or small, relatively to their variances. For this we set 0.6 and 0.3 as a covariance value for the case when the factor variances are large(0.8). And covariance values for the other cases are determined proportionally by rescaling Φ.

Hence, the number of conditions for Φ is $3 \times 2 = 6$; 3 for the values of factor variance, and 2 for those of factor covariance. These conditions are as follows.

$$\Phi_1 = \begin{pmatrix} 0.80 & 0.60 \\ 0.60 & 0.80 \end{pmatrix} \qquad \Phi_2 = \begin{pmatrix} 0.80 & 0.60 \\ 0.30 & 0.30 \end{pmatrix}$$

$$\Phi_3 = \begin{pmatrix} 0.60 & 0.45 \\ 0.45 & 0.60 \end{pmatrix} \qquad \Phi_4 = \begin{pmatrix} 0.60 & 0.225 \\ 0.225 & 0.60 \end{pmatrix}$$

$$\Phi_5 = \begin{pmatrix} 0.40 & 0.30 \\ 0.30 & 0.40 \end{pmatrix} \qquad \Phi_6 = \begin{pmatrix} 0.40 & 0.15 \\ 0.15 & 0.40 \end{pmatrix}$$

3) Specifying the Size of Model Error δ

In Section 5.2, we announced that several simulation studies will be conducted considering two cases; one for the case the model error is not concerned, and the other for the case we employ the model error in data-generating process. Then, how can we determine the size of model error δ? Theoretically, there is no correct answer to this question since the value of δ is unknown and cannot be studied in real-world data. Also, true degrees of model error are different depending on the phenomena researchers aim to study. Moreover there is no proper reference on this issue. There are only number of researches that concern the concept of model error in SEM and none of them deal with determining an appropriate size of δ.

As one of available approaches to solve this problem, we determined the value based on a moderate size of DE, or sampling error observed in the correctly specified model with no model error. In other words, the size of δ is given as equal to DE value of true model when the sample size is at least moderate. At first, we carried out simulations studies with no model error design. In these analyses, we could observe estimates of DE when the third and fourth condition of sample size is employed; (200, 1,000) for the factor analysis model, and (500, 1,000) for the structural equation modeling. In model 1 with no model error, DE and OD are almost equal to each other since there is no model error, or DA. Since we include 6 conditions on Φ matrix, $2 \times 6 = 12$ DE values were observed for both models.

Using this result, we determined δ as the mean of those 12 DE estimates. By this strategy, we obtained $\delta = 0.054728$ for the FA model, and $\delta = 0.052478$ for the SEM. Those mean DE estimates were not changed depending on the type of model used in each of simulations, and were similar with each other.

In their proposal for the model error inserting procedure, Cudeck and Browne added model error matrix E with $\delta = 0.25$ to their quasi-true covariance matrix $\Sigma_0(\theta_0)$ whose mean value of diagonal elements was about 4.5. Diagonal elements of $\Sigma_0(\theta_0)$ in the present simulation studies were all set to be 1, so the proportion of model error size and averaged value of variance elements were similar to the example in Cudeck-Browne procedure. Of course, this point was just one of our reference and cannot be a definite reason to support our determination.

The reason we exclude results from the first and second conditions of sample sizes, (50, 100) for FA and (150, 250) for SEM, is that sampling error, or DE, can be excessively amplified if n is small. If we included these cases, the values of δ were far larger than what we really used. This means that our theory and model were not appropriate for approximating the real-world phenomenon. Considering the definition of model error, it should be reduced by careful scrutiny, hypothesizing, and modeling of the interested phenomenon, not by statistical methods if the error is too large. Therefore, it is not worth studying the performance of estimation methods with this large scaled model error; if it is the case, researcher should turn back to their 'theorization' stage.

With our specification on δ, we expect the consequences for the model error case simulations as follows.

i) When we deal with the model 1 which has no misspecification, prespecified δ value will be recovered as a DA estimate when the ML estimation method is exploited.

ii) In the other models, where some misspecified parameters are added, DA values will be increased when the ML method is exploited. This is due to the fact that we use model 1 as a quasi-true model and generate pop-

ulation using this process with adding model error($\Sigma_0 = \Sigma_0(\theta_0) + \tilde{E}$). In other words, this model 1 should act as a best approximating model, and when several misspecifications are added, the approximation should be worsened.

iii) When Lasso is used in estimation, DA will be higher than ML's result in model 1. This is due to the Lasso's shrinkage estimation, which generates bias.

iv) In the other models, it is not easy to predict the result, since Lasso will also be suffer from misspecification but it can remove those unnecessary parameters. If Lasso is able to leave all the misspecification out, and degrees of shrinkage for correctly specified paths are not that large, we can expect that Lasso may be able to yield better DA.

4) Specifying Hyperparameters in Bayesian/BLasso SEM

In Bayesian Analysis, hyperparameters which define and control prior distributions can be important. Actually, previous researchers(Song & Lee, 2012; Guo et al., 2012) clarify the values of hyperparameters they use in analyzing their models, and also reveal that the results are not much affected when they examine various values of those parameters.

Based on their conclusion, we specified hyperparameters appropriately to the range of values of each parameter; *For all* j, k, *and* l, $\alpha_{0\epsilon j} = \alpha_{0\zeta k} = 25$, $\beta_{0\epsilon j} = \beta_{0\zeta k} = 7$, $\Lambda_{0jk} = \Lambda_{0\eta kl} = 0$, *and* $\rho_0 = 7$, $R_0 = \rho_0 \cdot \Phi^s$. *Also* $\alpha_{0\Lambda_j} = 4$, $\beta_{0\Lambda_j} = 7$ *are used in FA.* Moreover, posterior samples of size $4,000$ are obtained for each of iteration, a quarter of which are discarded as 'Burn-in' period. The remaining MCMC chain of size $3,000$ are used in computing several posterior estimates,

implied covariance matrices and fit indices.

BLasso SEM has another important issue – how can we determine the tuning parameter? Fortunately, Park and Casella(2008) suggest the use of hyperpriors for this parameter. With this approach, it can be expected that the optimum value of κ will determined automatically in the process of posterior estimation in BLasso. Therefore, we use this approach in the present FA simulation studies. SEM simulations were also carried out with this method, but a severe problem occurred; some posterior distributions were not converged in many cases. For example, some of the regression coefficients showed negative or too large estimates in most cases. Those results seemed too implausible considering their true values shown in Section 5.3. It is difficult to find the right reason to this phenomenon, but one possible cause is on the determined values of κ. With the hyperprior strategy, the posterior estimates of tuning parameters tend to be too small. Therefore, they cannot effectively control the corresponding coefficients, only to incur their non-convergence.

In this regard, the ordinary approach is exploited in the present SEM simulations. Tuning parameters can be determined by the same optimization procedure with Lasso(subsection 4.3.3) before BLasso SEM is fitted. However, BLasso SEM takes too much time per each iteration since they need to iterative sampling procedure during derivation of posterior distributions. Thus, it will be extremely tedious and time-consuming if we conduct this optimization for all iteration case.

Specifically, note that we consider 7 models(4 for FA, 3 for SEM), 4 sample size conditions, 6 Φ matrix conditions, 2 cases regarding model error(including error or not), and 100 repetition for each condition. This yields $7 \times 4 \times 6 \times 2 \times 100 = 33,600$ iterations. The optimization method of Lasso in subsection 4.3.3 says that we should set arbitrary t numbers of candidate values for κ, and

carry out Lasso estimation with each of them. This gives us t results and we can adopt one that yields the minimum valued criterion. With the optimization process of this kind, the number of total trial increases t times. If we have 30 candidates for κ, BLasso SEM should be carried out 30 times for each of iterations, which yields total 1,008,000 BLasso SEM fitting. Exactly the same thing occurs for L_1-regularized SEM, but it is not a big problem since Lasso SEM function spends much less time estimating than BLasso. On the contrary, BLasso SEM, which contains dragging updating and posterior sampling procedure, is vulnerable to this problem and its total simulation time can easily become impractical.

In this regard, we exploited another approach to obtain the optimum value for tuning parameter for BLasso SEM. At the outset, we obtained samples from the population-generating process, without sampling error; those samples have the same covariance matrix with our population regardless of their sample sizes. Sampling was done with different conditions on sample sizes and Φ matrices which is equal to what we use in main analysis. Therefore, the outcome can be seen as 'small-scale populations' for each of $4 \times 6 = 24$ conditions. Our approach is to determine κ by conducting the ordinary optimization procedure with these samples, not with every single sample obtained in each iteration of main analysis. Then the values are pre-determined beforehand, separately for each of conditions, and each iteration in the main BLasso SEM simulations used these values according to the sample sizes and Φ matrix conditions.

For this procedure, we drew the trace plots whose x-axis consists of candidate values of κ, and y-axis shows the values of optimization criteria – OD and MSE. What is the most unique in those plots is that they share the similar trend; as κ is getting larger, the values of criteria fell sharply at first, sometimes

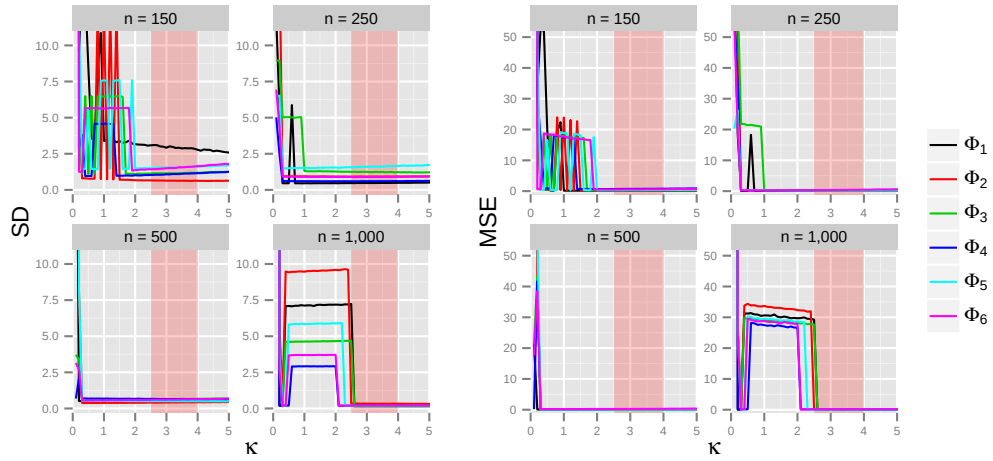

Figure 5.3: κ Trace Plots for BLasso SEM. Plots are drawn separated by sample sizes and Φ matrices conditions. Red-colored range indicates $2 \leq \kappa \leq 3.5$, where we determined values of the tuning parameters. Determined values are different depending on conditions.

with oscillation. After this rapid drop they become stabilized, mostly increasing only minimally. We interpreted this result as i) at first the estimation process finds the area close to the minimum as κ grows, reducing OD and MSE, and then ii) keep increasing κ causes the degree of shrinkage become excessive, aggravating those criteria. Hence, it seems to be sufficient to determine κ not too small, but not too large either.

With careful scrutiny on trace plots, we determined κ to be about $2 \sim 3.5$. After this preprocess, the main BLasso SEM simulations are conducted with these κ values. The results show proper convergences and estimation outcomes, unlike the former case we used hyperpriors for κ. The same procedure can be applied to population with DA as a criterion, in order to determine values of κ for population analysis. This lead us to the conclusion that $\kappa = 20$ is well-functioning for all Φ conditions.

We close this section by adding some comments on the issue. 'i) The optimization of κ can be done with the hyperprior strategy suggested by Park and Casella(2008), but it seems sometimes the method doesn't work for BLasso SEM. This result can be changed depending on several conditions – types of model, model complexity, sample sizes, and so on. ii) Furthermore, in practical cases where researchers normally have to deal with only a few datasets, the ordinary optimization method can also be applied to BLasso SEM without consuming too much time. Only t-times of estimation are needed for analyzing real data with BLasso SEM. With this precise optimization, the Bayesian approach for L_1-regularized SEM may produce better results than what we will show in the following chapter. Readers should be cautious about that our BLasso SEM results in the present thesis are not with full-optimization procedure – thus, the method can be better in practice.

5.5 Indices in Simulation Study

A variety of indices are computed during our simulation studies for several research purposes. These indices can be grouped as follows.

Matrix Discrepancy Discrepancy values, which are illustrated in Section 1.4 are computed. Among these variables are OD, DA, DE, and SD. Note that our main interest is focused on OD, which acts as an indices for generalizability of model. DE also receives huge attention since it contains a degree of variability of model. Since these indices are computed using covariance matrices, we propose to call these as 'Matrix Discrepancies'.

Each calculation can be done with various types of discrepancy functions such as ML, GLS, and OLS. Since ML and GLS function are prone to ML method, OLS is used in the present simulation studies as a fair criterion for our purpose. Outcomes with the other types of discrepancy function are also computed.

Parameters Discrepancy By the term 'parameter discrepancy' we indicate the indices related to MSE; that is, MSE, bias^2, and variance. These indices can be regarded as representing discrepancies between parameters, not their covariance matrix. This type of discrepancies have not been studied much in SEM. It seems that this is due to the fact that SEM is mainly interested in the overall model fit. However, in order to obtain a more generalizable and reproducible result, model specification and parameter estimation should receive more attention. In this regard, parameter discrepancy indices can be great criteria for indicating a good estimation method.

Out of parameter discrepancy, our main interest is MSE, which is a counter-

part of OD in matrix discrepancies. Low variance of a model is also accepted as a good property.

Also note that there are two type of estimate - unstandardized and standardized, in SEM. Thus, parameter discrepancies can be defined in two different ways. In the present thesis, MSE computed with unstandardized estimates is denoted as 'MSE'. And the other case, MSE computed with standardized estimates, is called as 'sMSE', which is abbreviation for 'MSE of Standardized Parameter Estimates'. In addition, we shall focus on sMSE since practical interpretation of model is carried out with standardized estimates, while unstandardized estimates are mainly used in estimation process.

In computation of parameter discrepancies, expected values of estimators are required. Note that it is equal to $\tilde{\theta}$ in figure 1.3, which is a vector containing the estimates obtained using all the population. Therefore, MSE, bias squared, and variance of estimators can be computed using $\tilde{\theta}$.

Fit Indices Various goodness of fit indices are also computed. They can be categorized as described in Section 1.3.

1) Absolute Fit indices : RMSEA, SRMR, Mc, $\hat{\gamma}$
2) Incremental Fit indices : NFI, TLI, CFI, RFI(=BL86), IFL(BL89), RNI
3) Information Criteria : AIC and BIC
4) Expected Cross Validation Index : ECVI

In section 1.3-1.4, we discussed that some of those indices can be useful for achieving more generalizable and reproducible models while others cannot. ECVI, AIC and BIC are included in the former group since these indices are defined in the context of cross validation and model selection. However, one thing should be noted; these indices work in a different way from the other

indices. Generally, they are used in model comparison and selection where the values of those indices do not matter. What is important is the order of values of AIC, BIC, or ECVI among various candidate models. The model producing the smallest values of these indices is adopted in model selection. Thus, a value itself has no information in this regard. Therefore, relationships between these criteria and discrepancy indices should be studied again, in a different design.

Incremental fit indices are expected not to be good enough for our purpose due to the fact that the they are heavily related to SD, which is severely affected by the current sample. Similar comments can be made regarding absolute fit indices. However, some of them, such as RMSEA and Mc which are related to DA, seem to perform in a different way.

In the present simulation studies, discrepancy and parameter discrepancy indices are mainly used in comparison of estimation methods including ML, Lasso, and BLasso. However, since they are not observable in practical sample data, fit indices are computed and correlations among all the indices are calculated. This aims to figure out which of those fit indices show strong relationships with generalizability indices like OD and MSE. Basically, investigation on the relationships should be conducted with studies on mathematical properties of each index. However, in the present thesis, we carry out this correlation study to comprehend the empirical features and tendencies of those indices.

As matrix discrepancies, fit indices can be computed with different discrepancy functions. We obtain and scrutinize all outcomes.

5.6 Flow of Simulation

For both of the models, simulation is conducted according to the following order.

Step 1. Generate Σ_0.
With predetermined quasi-true model and prespecified parameters, compute $\Sigma_0(\theta_0)$. Also add a perturbation matrix E in accordance with Cudeck and Browne(1992)'s procedure. At last, obtain $\Sigma_0 = \Sigma_0(\theta_0) + E$. The result matrix varies depending on the condition of Φ illustrated in Section 5.4.

Step 2. Generate the population.
Generate the population data whose covariance matrix is equal to Σ_0. The size of population is determined to be $N = 50,000$.

Step 3. Analyze the population.
Fitting the CFA/SEM models to the population data by ML, Lasso, and BLasso. The models are illustrated in Section 5.3. From the result, we can obtain DA and bias^2.

Step 4. Sampling.
Sample the data from the population generated in **Step 2**. The sample size varies according to the condition described in Section 5.4.

Step 5. Analyze the sample.
Fitting the CFA/SEM models to the sample data by ML, Lasso, BLasso. From the result we can obtain SD and various fit indices. Also, OD, DE, MSE, and variance can also be computed using the result from **Step 2** and this step.

Step 6. Repeat under the randomness.

Repeat **Step 4-5**, 100 times. With 100 estimates for each parameter, we can investigate and compare performances of ML, Lasso, and BLasso.

Step 7. Repeat under the sample size condition.

Repeat **Step 4-6** with another condition of sample size. Conduct this step until all of the sample size conditions are completed.

Step 8. Repeat under the Φ condition.

Repeat **Step 1-7** with another condition of Φ. Conduct this step until all of the Φ conditions are completed.

Step 9. Repeat with another model.

Repeat the total simulation (**Step 1-8**) with another model.

Chapter 6
Simulation Study : Result

This chapter presents the results of simulations studies. Research 1 studies on the performance of each estimation methods when the factor analysis model is fitted, while Research 2 investigates the same topics for the structural equation modeling. Note that for each of models, $4 \times 6 = 24$ conditions are taken into consideration, including various sample sizes and Φ matrices. In order to indicate each condition, we use the expression '*Condition* (i, j)' where i denotes the condition of sample size and j denotes that of Φ. For example, $(2, 3)$ indicates the case when the second condition of sample size condition(100 for FA, and 250 for SEM) and the matrix Φ_3 are considered in the simulation.

6.1 Research 1: Factor Analysis Model

Figure 6.1 and 6.2 display the density plots of OD obtained by BLasso for model 3, one for the case when the model error is not considered and the other for when the prespecified model error is involved. This densities are computed using 100 iterated simulation results. The x-axis is adjusted to be $[0, 1.25]$ in order to represent the effect of increasing sample sizes on OD distribution.

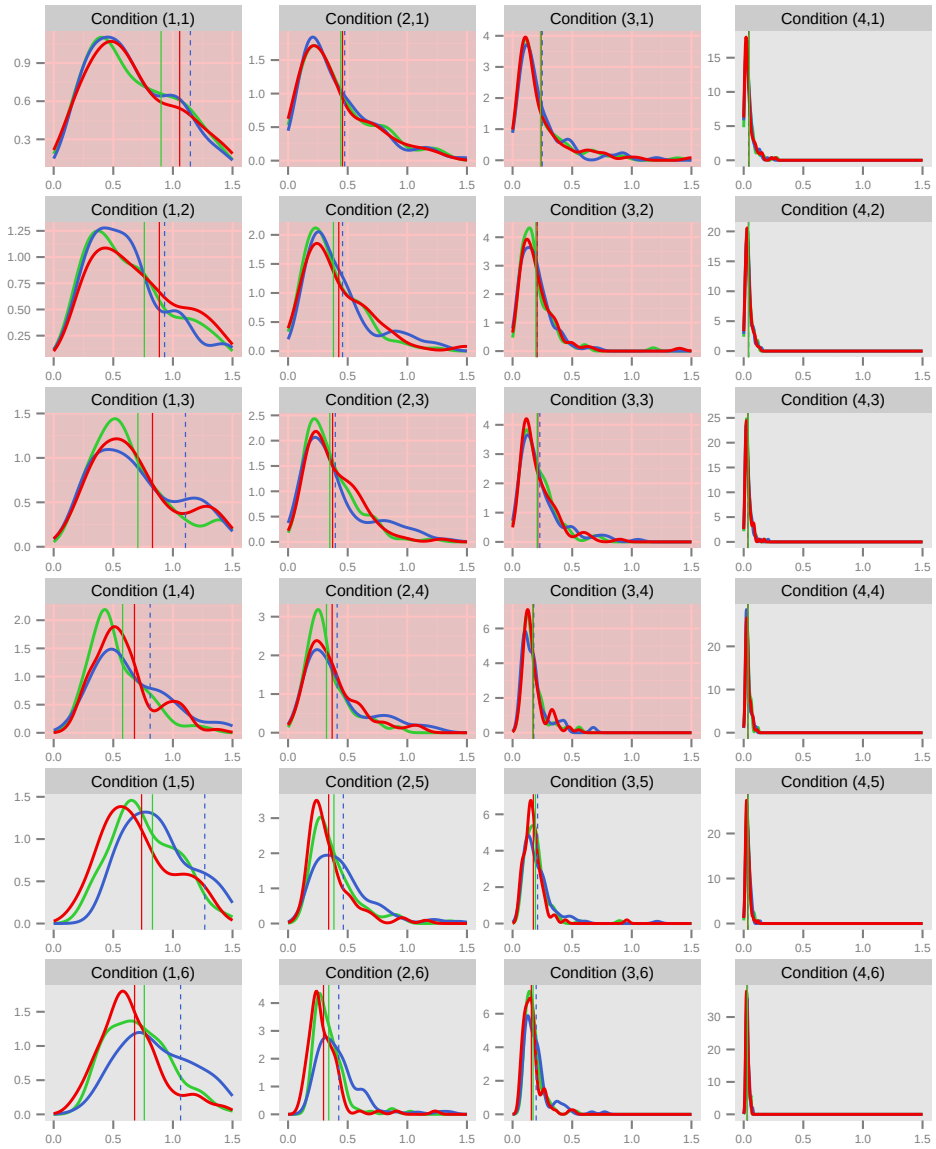

Figure 6.1: Density Plots of Overall Discrepancy - BLasso, No Model Error

In the plots, red, blue, and green lines indicate ML, posterior mean of BLasso, and MAP of BLasso, respectively. And the vertical lines represent the averaged values of 100 OD observations. These are the estimates of parameter OD in

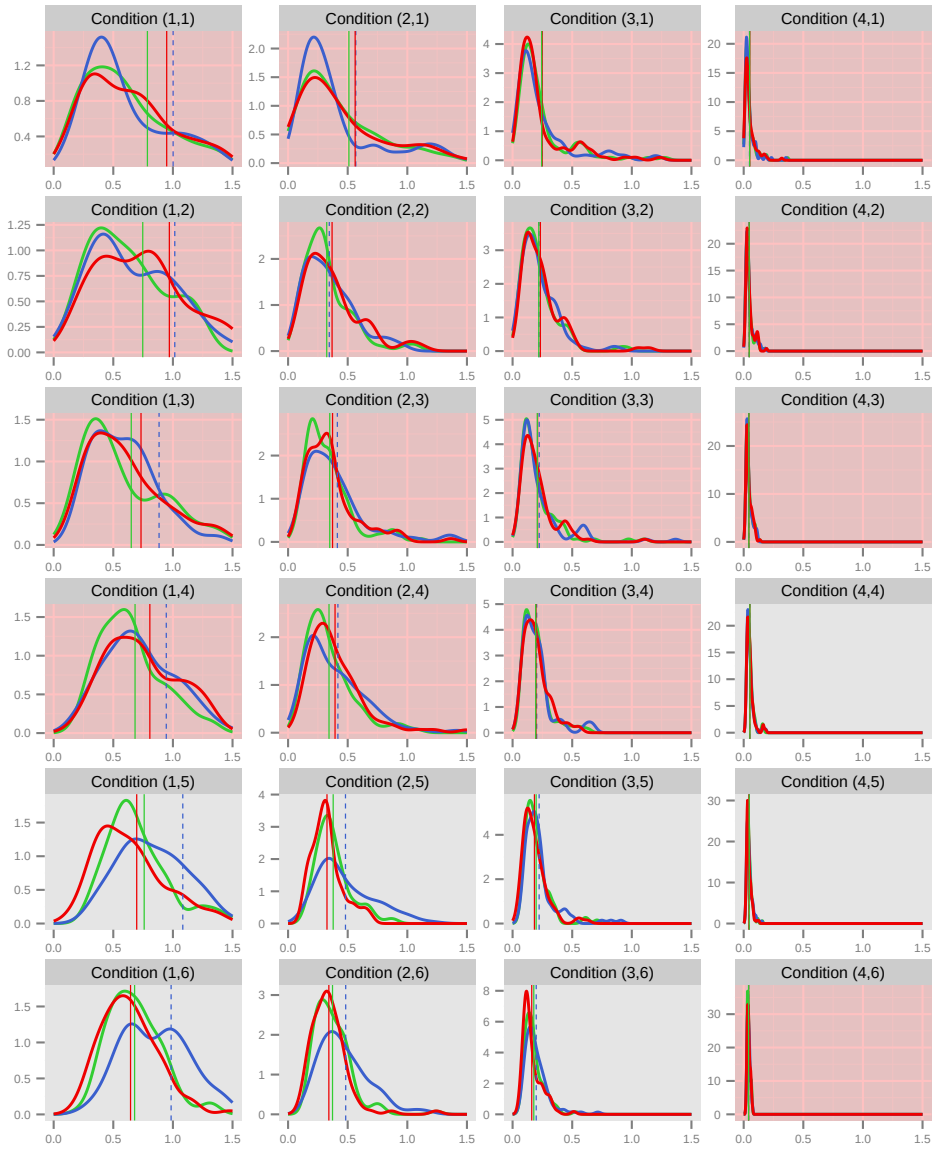

Figure 6.2: Density Plots of Overall Discrepancy - BLasso, Model Error Involved

the corresponding conditions. For BLasso FA, MAP outperforms the posterior mean in producing less mean values of OD. Therefore, we focus on comparing ML and MAP of BLasso, with presenting the result of posterior mean together

for information. For representing this point, the mean OD line of posterior mean is displayed as a dotted line.

In conditions where the mean OD of BLasso is less than that of ML, the corresponding panels are colored light red. Other models yield similar results, except model 4 which shows the less supportive result for BLasso. Since the purpose of investigating BLasso is on the question asking whether the method is able to produce complete shrinkage, not on the OD estimates, only the model 3 result is suggested as a representative case. Instead, mean OD values computed by BLasso for the other models are posted in Appendix A. The figures imply that BLasso is able to produce less OD than ML in several conditions. And the tendency is quite systematic; For $\Phi_1 \sim \Phi_4$, BLasso outperforms ML whereas the opposite occurs for Φ_5 and Φ_6. When the sample size is large enough, namely, 1,000, ML shows greate result by the virtue of its asymptotic properties. Also note that when the model error is involved, it aggravate ML's results which leads BLasso to produce better performance.

Table 6.1 and 6.2 imply the answer to the question on the shrinkage of BLasso. In those table, 'Total Number of Trials' represents the number of occurrences of parameters, which are supposed to be removed by BLasso, over all the 4×6 conditions and 100 iterations. For example, four misspecified parameters($\gamma_3 - \gamma_6$) are included in model 3. Therefore, $4 \times 6 \times 100 \times 4 = 9,600$. However, in some iterations ML produces non-convergence, which are excluded from the count. Subtracting the number of non-convergence, 'Number of Shrinkage Cases' are calculated for each of models.

Since it is known from Park and Casella(2008) that BLasso cannot produce the exact-zero shrinkage, we test the strategy that considers the BLasso estimates as zero when they are less than the pre-determined bound. For this purpose, we set investigate a string of bounds from 10^{-1} to 10^{-7}. It should

Shrinakge Effect in BLasso FA : Without Model Error

	Model 2				*Model 3*				*Model 4*			
	Total Number of Trials		2,400		Total Number of Trials		9,600		Total Number of Trials		2,400	
	Number of Non-convergence		4		Number of Non-convergence		112		Number of Non-convergence		-	
	Number of Shrinkage Cases		2,396		Number of Shrinkage Cases		9,488		Number of Shrinkage Cases		2,400	
Bound	ML		BLasso		ML		BLasso		ML		BLasso	
10^{-1}	1,493	(62.3%)	1,710	(71.4%)	5,612	(59.0%)	6,412	(67.6%)	1,486	(61.9%)	1,620	(67.5%)
10^{-2}	230	(9.60%)	333	(13.9%)	837	(8.8%)	1,095	(11.5%)	181	(7.54%)	244	(10.2%)
10^{-3}	26	(1.09%)	31	(1.29%)	92	(0.97%)	97	(1.02%)	11	(0.46%)	30	(1.25%)
10^{-4}	1	(0.04%)	-	-	10	(0.11%)	4	(0.04%)	1	(0.04%)	4	(0.17%)
10^{-5}	-	-	-	-	-	-	1	(0.01%)	-	-	-	-
10^{-6}	-	-	-	-	-	-	1	(0.01%)	-	-	-	-
10^{-7}	-	-	-	-	-	-	-	-	-	-	-	-

Table 6.1: Shrinkage Effect in BLasso FA - Without Model Error

Shrinakge Effect in BLasso FA : With Model Error δ

	Model 2				*Model 3*				*Model 4*			
	Total Number of Trials		2,400		Total Number of Trials		9,600		Total Number of Trials		2,400	
	Number of Non-convergence		1		Number of Non-convergence		52		Number of Non-convergence		1	
	Number of Shrinkage Cases		2,399		Number of Shrinkage Cases		9,548		Number of Shrinkage Cases		2,399	
Bound	ML		BLasso		ML		BLasso		ML		BLasso	
10^{-1}	1,211	(50.5%)	1,471	(61.3%)	3,631	(38.0%)	4,449	(46.6%)	1,526	(63.6%)	1,644	(68.5%)
10^{-2}	97	(4.04%)	220	(9.17%)	372	(3.90%)	630	(6.60%)	227	(9.5%)	283	(11.8%)
10^{-3}	13	(0.54%)	21	(0.88%)	43	(0.45%)	65	(0.68%)	22	(0.92%)	29	(1.21%)
10^{-4}	-	-	3	(0.13%)	5	(0.05%)	7	(0.07%)	-	-	4	(0.17%)
10^{-5}	-	-	3	(0.13%)	-	-	-	-	-	-	-	-
10^{-6}	-	-	1	(0.04%)	-	-	-	-	-	-	-	-
10^{-7}	-	-	1	(0.04%)	-	-	-	-	-	-	-	-

Table 6.2: Shrinkage Effect in BLasso FA - With Model Error δ

be noted that the same approach can be applied to ML results. Therefore, we also test ML with BLasso. If ML yields as many cases of removing parame-

ters as BLasso by this strategy, it cannot be said that BLasso has a capacity of shrinking the coefficient completely to zero.

The result implies that the strategy doesn't work properly. Using 10^{-1} as a

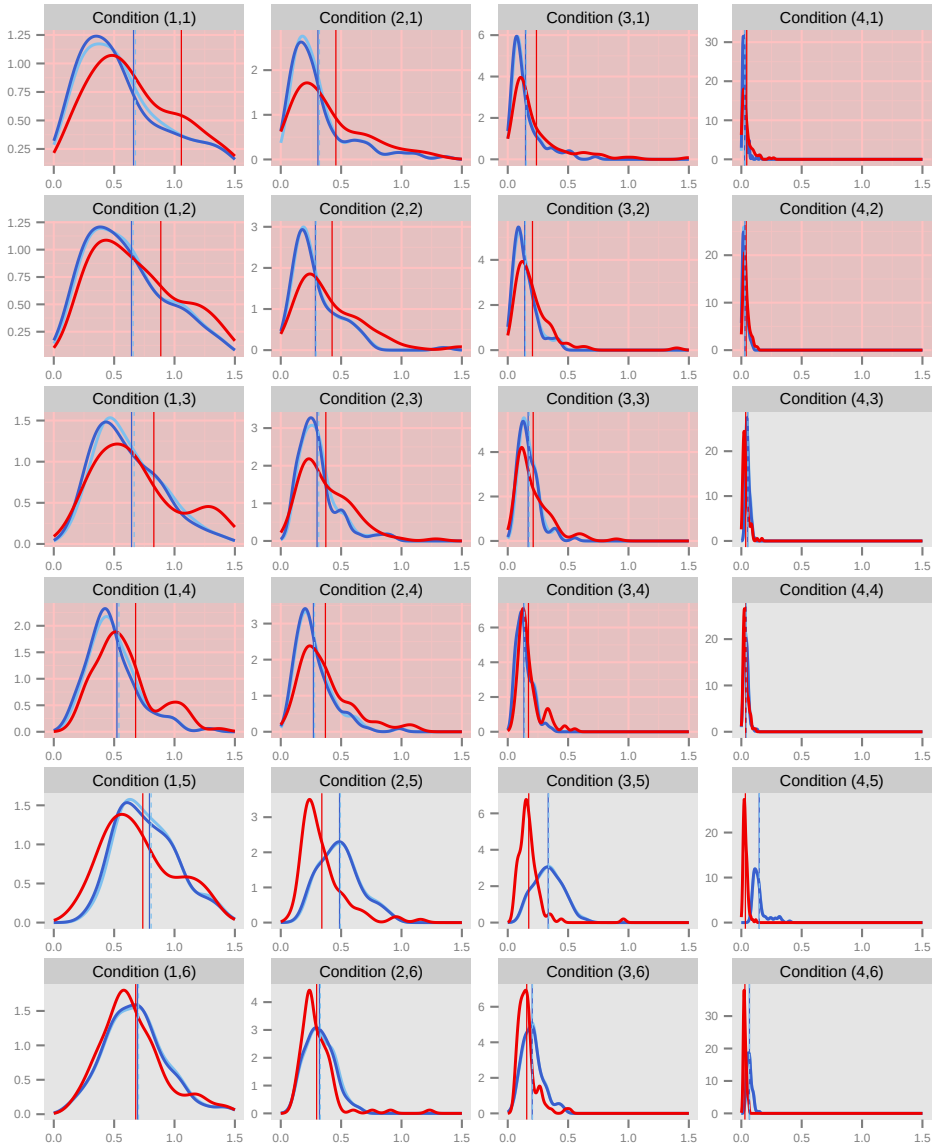

Figure 6.3: Density Plots of Overall Discrepancy - Lasso FA, No Model Error

bound, we may regard many of the cases as being shrunken to zero by BLasso. However, this bound is not appropriate since it is not small enough; note that normally at least about 10^{-6} is used as this kind of bound, and also note that

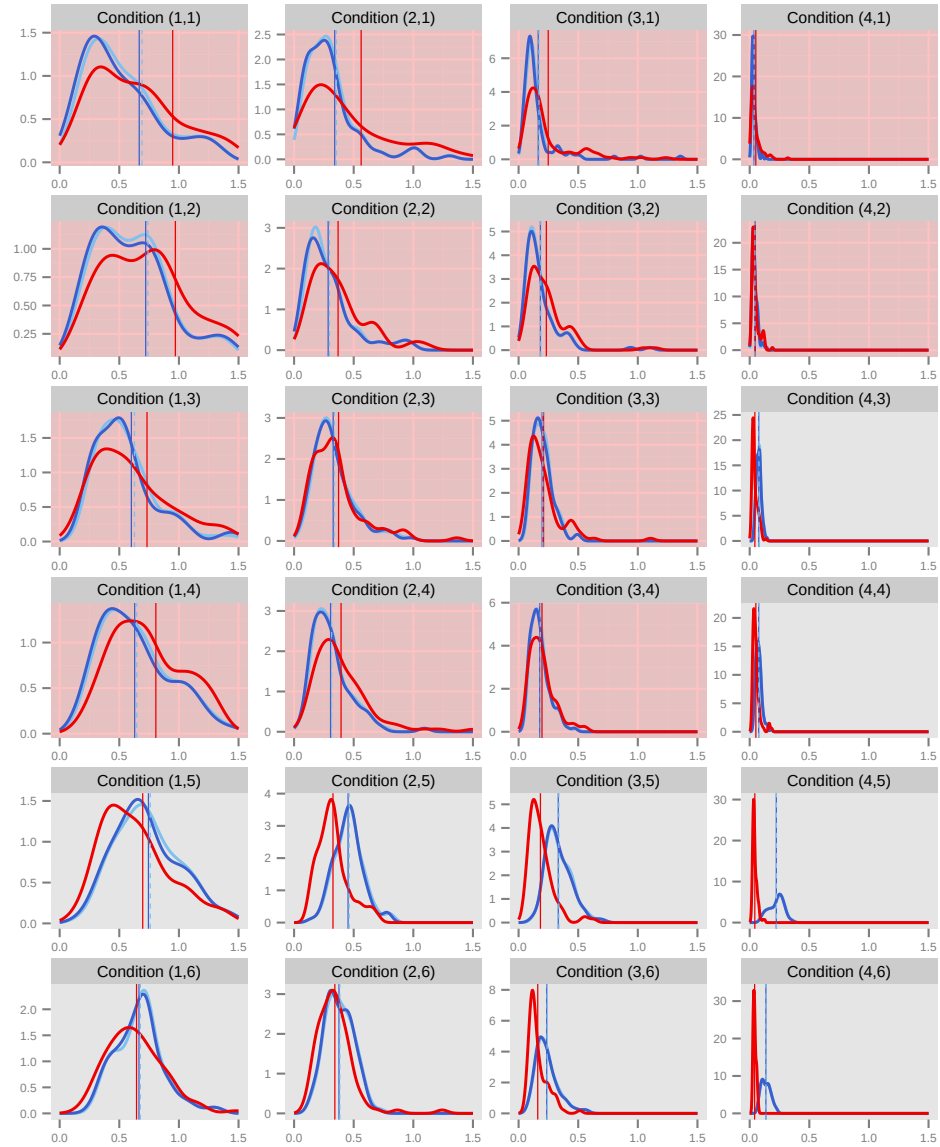

Figure 6.4: Density Plots of Overall Discrepancy - Lasso FA, Model Error Involved

we use 10^{-15} for Lasso result. Moreover, it cannot be said that BLasso is able to delete some coefficients when the bound is less than 10^{-2}. More importantly, ML also produce similar results with those of BLasso. With this strategy, ML and BLasso differ only slightly and do not show any remarkable difference.

Next, results of the proposed method, Lasso FA, are provided. Figure 6.3 and 6.4 contains the OD density plots as 6.1 and 6.2, but for Lasso FA. Note that as in the BLasso FA, we only present the model 3 case as in the BLasso FA. Additionally, averaged OD values for all the conditions and models are suggested in table 6.3 and 6.4, for the case without and with model error involved, respectively.

In the figure, red line indicates the result of ML and blue and skyblue lines represent Lasso cases. Difference between those two Lasso results are in the tuning criterion used in optimization; the former, which we denote as 'Lasso.OD', assumes that we can approach the true population covariance matrix so that the true OD values are computed and used, and the latter, indicated as 'Lasso', doesn't lean on this assumption so that Lasso results are optimized by means of the cross-validation method. For the purpose of investigating the potentially best performance of Lasso, the comparison will be based on ML and Lasso.OD. And by scrutinizing Lasso result, we can test whether the proposed cross-validation method works well. Note that optimization based on $SD.cv$, our cross-validation criterion, yields the very close result to that of the case when the OD is used.

Based on the estimates of OD, it turns out that Lasso can produce better result than ML in the case 1) the Φ matrix is set to be Φ_1 or Φ_2, 2) sample size is not too large when $\Phi = \Phi_3, \Phi_4$. On the contrary, ML outperforms when the Φ matrix condition is Φ_5 or Φ_6. Also, for Φ_3 and Φ_4, ML yields great result when the sample size is large, namely 1,000. Therefore, performances of those

OD : Overall Discrepancy (OLS)

Sample Size	Φ	Model 1 ML	Model 1 Lasso	Model 1 Lasso (OD)	Model 2 ML	Model 2 Lasso	Model 2 Lasso (OD)
50	Φ_1	0.9443	0.6400 (0.85)	0.6156 (0.96)	1.0040	0.7259 (0.87)	0.6955 (0.94)
	Φ_2	0.7676	0.5877 (0.86)	0.5710 (0.93)	0.7378	0.5599 (0.81)	0.5400 (0.94)
	Φ_3	0.7939	0.6168 (0.67)	0.5963 (0.74)	0.7790	0.5594 (0.66)	0.5363 (0.71)
	Φ_4	0.6764	0.5018 (0.75)	0.4871 (0.77)	0.6597	0.5514 (0.62)	0.5359 (0.66)
	Φ_5	0.6738	0.7191 (0.20)	0.7025 (0.24)	0.5733	0.7309 (0.20)	0.7180 (0.21)
	Φ_6	0.5401	0.6034 (0.28)	0.5925 (0.30)	0.5333	0.5557 (0.31)	0.5472 (0.32)
100	Φ_1	0.3873	0.2809 (0.78)	0.2707 (0.87)	0.4576	0.3149 (0.81)	0.2970 (0.94)
	Φ_2	0.3551	0.2804 (0.78)	0.2705 (0.84)	0.3765	0.2942 (0.80)	0.2878 (0.83)
	Φ_3	0.3950	0.2915 (0.61)	0.2787 (0.68)	0.3374	0.2984 (0.51)	0.2811 (0.58)
	Φ_4	0.3391	0.2582 (0.61)	0.2488 (0.67)	0.2972	0.2375 (0.60)	0.2283 (0.66)
	Φ_5	0.2872	0.4115 (0.14)	0.4037 (0.15)	0.3091	0.4391 (0.16)	0.4309 (0.16)
	Φ_6	0.2748	0.3395 (0.25)	0.3330 (0.27)	0.2889	0.3548 (0.23)	0.3476 (0.24)
200	Φ_1	0.2329	0.1469 (0.80)	0.1392 (0.88)	0.2222	0.1317 (0.79)	0.1225 (0.88)
	Φ_2	0.2090	0.1639 (0.78)	0.1595 (0.83)	0.2020	0.1592 (0.76)	0.1553 (0.82)
	Φ_3	0.1653	0.1584 (0.38)	0.1491 (0.43)	0.1860	0.1632 (0.49)	0.1579 (0.49)
	Φ_4	0.1315	0.1241 (0.49)	0.1185 (0.56)	0.1815	0.1510 (0.59)	0.1459 (0.66)
	Φ_5	0.1604	0.2681 (0.12)	0.2637 (0.13)	0.1584	0.2823 (0.11)	0.2774 (0.11)
	Φ_6	0.1326	0.1974 (0.17)	0.1952 (0.17)	0.1373	0.1941 (0.19)	0.1924 (0.20)
1000	Φ_1	0.0475	0.0289 (0.63)	0.0271 (0.69)	0.0440	0.0306 (0.62)	0.0287 (0.66)
	Φ_2	0.0316	0.0255 (0.63)	0.0242 (0.71)	0.0354	0.0278 (0.65)	0.0266 (0.68)
	Φ_3	0.0360	0.0607 (0.11)	0.0587 (0.11)	0.0307	0.0601 (0.11)	0.0584 (0.13)
	Φ_4	0.0310	0.0411 (0.21)	0.0401 (0.21)	0.0305	0.0401 (0.21)	0.0389 (0.22)
	Φ_5	0.0294	0.1302 (0.01)	0.1293 (0.01)	0.0327	0.1349 (0.02)	0.1341 (0.02)
	Φ_6	0.0254	0.0714 (0.03)	0.0709 (0.03)	0.0273	0.0660 (0.05)	0.0655 (0.07)

Sample Size	Φ	Model 3 ML	Model 3 Lasso	Model 3 Lasso (OD)	Model 4 ML	Model 4 Lasso	Model 4 Lasso (OD)
50	Φ_1	1.0567	0.6762 (0.88)	0.6606 (0.92)	1.0787	0.6506 (0.87)	0.6274 (0.96)
	Φ_2	0.8863	0.6571 (0.87)	0.6436 (0.96)	0.7471	0.5958 (0.88)	0.5807 (0.93)
	Φ_3	0.8288	0.6658 (0.54)	0.6442 (0.57)	0.9087	0.6235 (0.69)	0.6028 (0.71)
	Φ_4	0.6776	0.5394 (0.66)	0.5229 (0.72)	0.6764	0.5514 (0.75)	0.5392 (0.78)
	Φ_5	0.7372	0.8065 (0.26)	0.7925 (0.28)	0.6997	0.8062 (0.27)	0.7926 (0.28)
	Φ_6	0.6785	0.7018 (0.37)	0.6935 (0.38)	0.6566	0.6641 (0.35)	0.6516 (0.35)
100	Φ_1	0.4565	0.3199 (0.79)	0.3079 (0.91)	0.5584	0.3455 (0.88)	0.3274 (0.96)
	Φ_2	0.4251	0.2924 (0.92)	0.2871 (0.94)	0.4210	0.2950 (0.85)	0.2867 (0.90)
	Φ_3	0.3734	0.3144 (0.63)	0.3005 (0.66)	0.4489	0.3263 (0.68)	0.3113 (0.74)
	Φ_4	0.3696	0.2797 (0.73)	0.2706 (0.80)	0.3504	0.2626 (0.78)	0.2524 (0.83)
	Φ_5	0.3402	0.4938 (0.14)	0.4869 (0.14)	0.3438	0.4379 (0.19)	0.4298 (0.19)
	Φ_6	0.2973	0.3279 (0.34)	0.3206 (0.36)	0.3018	0.3402 (0.33)	0.3342 (0.35)
200	Φ_1	0.2379	0.1530 (0.82)	0.1472 (0.90)	0.3010	0.1851 (0.89)	0.1783 (0.94)
	Φ_2	0.2048	0.1468 (0.85)	0.1403 (0.89)	0.2213	0.1528 (0.84)	0.1481 (0.90)
	Φ_3	0.2109	0.1781 (0.50)	0.1690 (0.54)	0.2283	0.1696 (0.66)	0.1629 (0.69)
	Φ_4	0.1716	0.1391 (0.67)	0.1339 (0.72)	0.1933	0.1446 (0.70)	0.1404 (0.72)
	Φ_5	0.1739	0.3392 (0.05)	0.3351 (0.06)	0.1599	0.2614 (0.10)	0.2577 (0.11)
	Φ_6	0.1572	0.2049 (0.21)	0.2028 (0.23)	0.1677	0.2029 (0.33)	0.1989 (0.35)
1000	Φ_1	0.0449	0.0290 (0.68)	0.0270 (0.76)	0.0417	0.0292 (0.60)	0.0263 (0.69)
	Φ_2	0.0398	0.0307 (0.68)	0.0291 (0.72)	0.0460	0.0310 (0.76)	0.0300 (0.78)
	Φ_3	0.0363	0.0546 (0.18)	0.0526 (0.18)	0.0433	0.0547 (0.23)	0.0529 (0.24)
	Φ_4	0.0356	0.0407 (0.33)	0.0396 (0.35)	0.0354	0.0392 (0.34)	0.0380 (0.37)
	Φ_5	0.0334	0.1490 (0.01)	0.1481 (0.01)	0.0303	0.1322 (0.00)	0.1313 (0.00)
	Φ_6	0.0295	0.0678 (0.04)	0.0671 (0.04)	0.0306	0.0726 (0.03)	0.0720 (0.03)

Table 6.3: Overall Discrepancy Table - Lasso FA, No Model Error

OD : Overall Discrepancy (OLS)

Sample Size	Φ	Model 1 ML	Model 1 Lasso	Model 1 Lasso (OD)	Model 2 ML	Model 2 Lasso	Model 2 Lasso (OD)
50	Φ_1	1.0726	0.7175 (0.81)	0.6973 (0.85)	0.9532	0.6501 (0.81)	0.6200 (0.87)
	Φ_2	0.8783	0.5959 (0.86)	0.5807 (0.93)	0.8639	0.6113 (0.88)	0.6002 (0.93)
	Φ_3	0.7632	0.6655 (0.54)	0.6478 (0.58)	0.7784	0.6685 (0.53)	0.6443 (0.57)
	Φ_4	0.6865	0.6149 (0.62)	0.5963 (0.70)	0.7361	0.5949 (0.69)	0.5803 (0.76)
	Φ_5	0.6796	0.8019 (0.24)	0.7896 (0.25)	0.6495	0.8291 (0.16)	0.8157 (0.16)
	Φ_6	0.6207	0.6805 (0.31)	0.6708 (0.32)	0.6611	0.7249 (0.34)	0.7151 (0.35)
100	Φ_1	0.4735	0.3550 (0.79)	0.3386 (0.91)	0.4719	0.3244 (0.78)	0.3137 (0.89)
	Φ_2	0.3907	0.3150 (0.80)	0.3052 (0.87)	0.4100	0.3167 (0.79)	0.3090 (0.89)
	Φ_3	0.3945	0.3581 (0.43)	0.3430 (0.46)	0.4091	0.3682 (0.44)	0.3564 (0.46)
	Φ_4	0.4256	0.3913 (0.48)	0.3847 (0.48)	0.3608	0.3178 (0.61)	0.3101 (0.62)
	Φ_5	0.3350	0.4962 (0.13)	0.4887 (0.13)	0.3514	0.5685 (0.10)	0.5629 (0.10)
	Φ_6	0.3188	0.4189 (0.22)	0.4135 (0.22)	0.3281	0.4403 (0.16)	0.4344 (0.17)
200	Φ_1	0.2282	0.1618 (0.72)	0.1529 (0.84)	0.2497	0.1772 (0.70)	0.1686 (0.75)
	Φ_2	0.2323	0.1673 (0.70)	0.1629 (0.76)	0.2093	0.1658 (0.74)	0.1602 (0.82)
	Φ_3	0.2135	0.2211 (0.32)	0.2148 (0.33)	0.2197	0.2293 (0.27)	0.2231 (0.31)
	Φ_4	0.2205	0.2140 (0.42)	0.2091 (0.43)	0.2110	0.2204 (0.29)	0.2137 (0.33)
	Φ_5	0.1727	0.3312 (0.05)	0.3268 (0.05)	0.1744	0.4082 (0.01)	0.4034 (0.01)
	Φ_6	0.1860	0.2942 (0.07)	0.2920 (0.07)	0.2154	0.3194 (0.13)	0.3163 (0.15)
1000	Φ_1	0.0549	0.0419 (0.51)	0.0406 (0.55)	0.0521	0.0407 (0.51)	0.0386 (0.58)
	Φ_2	0.0635	0.0557 (0.55)	0.0546 (0.58)	0.0600	0.0526 (0.58)	0.0513 (0.61)
	Φ_3	0.0625	0.1197 (0.03)	0.1181 (0.03)	0.0612	0.1251 (0.03)	0.1232 (0.03)
	Φ_4	0.0777	0.1184 (0.02)	0.1172 (0.02)	0.0783	0.1195 (0.03)	0.1185 (0.03)
	Φ_5	0.0636	0.2467 (0.00)	0.2458 (0.00)	0.0624	0.2848 (0.00)	0.2839 (0.00)
	Φ_6	0.0815	0.1768 (0.00)	0.1762 (0.00)	0.0851	0.1841 (0.01)	0.1835 (0.01)

Sample Size	Φ	Model 3 ML	Model 3 Lasso	Model 3 Lasso (OD)	Model 4 ML	Model 4 Lasso	Model 4 Lasso (OD)
50	Φ_1	0.9478	0.6903 (0.88)	0.6672 (0.93)	1.1193	0.7522 (0.88)	0.7245 (0.96)
	Φ_2	0.9701	0.7406 (0.90)	0.7213 (0.97)	0.8930	0.7046 (0.93)	0.6845 (0.96)
	Φ_3	0.7324	0.6273 (0.55)	0.6013 (0.61)	0.8984	0.7665 (0.60)	0.7509 (0.63)
	Φ_4	0.8066	0.6462 (0.73)	0.6289 (0.77)	0.7597	0.6647 (0.59)	0.6469 (0.68)
	Φ_5	0.6964	0.7597 (0.28)	0.7444 (0.29)	0.7423	0.9106 (0.22)	0.9014 (0.22)
	Φ_6	0.6446	0.6766 (0.31)	0.6650 (0.34)	0.7030	0.7217 (0.39)	0.7154 (0.39)
100	Φ_1	0.5624	0.3544 (0.79)	0.3406 (0.91)	0.5781	0.4008 (0.91)	0.3876 (0.95)
	Φ_2	0.3690	0.2963 (0.84)	0.2859 (0.95)	0.4554	0.3629 (0.89)	0.3542 (0.94)
	Φ_3	0.3730	0.3378 (0.46)	0.3280 (0.52)	0.4855	0.4120 (0.48)	0.3999 (0.48)
	Φ_4	0.3939	0.3146 (0.68)	0.3061 (0.74)	0.4205	0.3709 (0.62)	0.3602 (0.69)
	Φ_5	0.3263	0.4621 (0.11)	0.4529 (0.11)	0.3757	0.5266 (0.16)	0.5188 (0.18)
	Φ_6	0.3418	0.3839 (0.28)	0.3765 (0.28)	0.3840	0.4610 (0.26)	0.4566 (0.26)
200	Φ_1	0.2485	0.1713 (0.73)	0.1648 (0.84)	0.3069	0.2092 (0.86)	0.2028 (0.93)
	Φ_2	0.2329	0.1863 (0.79)	0.1829 (0.85)	0.2584	0.1941 (0.89)	0.1903 (0.92)
	Φ_3	0.2095	0.2034 (0.43)	0.1974 (0.46)	0.2320	0.2272 (0.42)	0.2194 (0.42)
	Φ_4	0.1964	0.1823 (0.57)	0.1769 (0.60)	0.2255	0.2259 (0.40)	0.2197 (0.46)
	Φ_5	0.1832	0.3378 (0.09)	0.3327 (0.10)	0.2088	0.3800 (0.11)	0.3752 (0.12)
	Φ_6	0.1605	0.2393 (0.14)	0.2360 (0.15)	0.2069	0.2902 (0.16)	0.2874 (0.16)
1000	Φ_1	0.0514	0.0373 (0.56)	0.0356 (0.59)	0.0553	0.0445 (0.58)	0.0423 (0.64)
	Φ_2	0.0471	0.0437 (0.46)	0.0423 (0.48)	0.0633	0.0567 (0.55)	0.0557 (0.59)
	Φ_3	0.0432	0.0778 (0.13)	0.0760 (0.14)	0.0664	0.1160 (0.05)	0.1144 (0.05)
	Φ_4	0.0518	0.0761 (0.13)	0.0753 (0.13)	0.0803	0.1164 (0.04)	0.1153 (0.04)
	Φ_5	0.0430	0.2243 (0.00)	0.2231 (0.00)	0.0672	0.2455 (0.00)	0.2447 (0.00)
	Φ_6	0.0413	0.1363 (0.00)	0.1357 (0.00)	0.0833	0.1727 (0.00)	0.1720 (0.00)

Table 6.4: Overall Discrepancy Table - Lasso FA, Model Error Involved

methods are affected by magnitudes of true population covariance matrix and sample sizes. In general, this tendency is similar with the BLasso result we investigated above.

In the table 6.3 and 6.4, the values of OD estimates are posted for ML, Lasso, and Lasso.OD, which indicates Lasso result with optimization based on OD. Comparing ML and Lasso.OD, the cells including less OD estimates are colored red. Lasso cells are also colored when it yields small estimates than ML, but in different color. Plus, note that columns corresponding to the Lasso results consist of two sub-columns; one for the estimates of OD described above, and the other for the rates which indicate the proportion Lasso and Lasso.OD results outperform that of ML among the 100 iteration. These 'rate cells' are also colored as similar tones with the cells containing estimates of OD. Color intensity depends on the rate; cells with $0.5 \sim 0.59$ has the same color with estimates cells, and cells with rate higher than 0.60 are colored with more intense tone. Also, cells with $0.40 \sim 0.49$ rate are colored with pale tone to represent the case Lasso does not outperform, but yields similar level of performance with ML.

The overall trend is similar with what we described above with figures. It should be noted that the trend is consistent over all the models. This is against our expectation that Lasso performs better as degree of misspecification increases. It seems misspecifications in Model 2, 3, and 4 do not have considerable effects on the estimation methods' performances. Furthermore, the tendency does not change according to whether the model error is involved or not, though the error slightly diminishes the difference of mean OD and outperformance rate. This result implies that the difference between ML and Lasso depends greatly on the conditions considered in the simulation - magnitudes of population covariances and the sample sizes. Also two differ-

sMSE : Mean Squared Error of Standardized Parameter Estimates

Sample Size	Φ	Model 1 ML	Model 1 Lasso	Model 1 Lasso (OD)	Model 2 ML	Model 2 Lasso	Model 2 Lasso (OD)
50	Φ_1	0.0687	0.0752 (0.51)	0.0720 (0.53)	0.0935	0.0928 (0.49)	0.0897 (0.46)
	Φ_2	0.0808	0.0871 (0.60)	0.0850 (0.62)	0.0945	0.0907 (0.66)	0.0878 (0.73)
	Φ_3	0.1898	0.2388 (0.16)	0.2289 (0.16)	0.2453	0.2593 (0.32)	0.2570 (0.32)
	Φ_4	0.2250	0.2584 (0.25)	0.2539 (0.23)	0.2537	0.2936 (0.29)	0.2866 (0.30)
	Φ_5	0.3476	0.7183 (0.00)	0.7034 (0.01)	1.9992	2.0759 (0.19)	2.2161 (0.21)
	Φ_6	0.4925	0.7697 (0.11)	0.7483 (0.12)	0.4783	0.7100 (0.09)	0.7091 (0.09)
100	Φ_1	0.0302	0.0310 (0.44)	0.0306 (0.42)	0.0462	0.0451 (0.49)	0.0435 (0.50)
	Φ_2	0.0400	0.0405 (0.54)	0.0402 (0.60)	0.0481	0.0479 (0.55)	0.0464 (0.61)
	Φ_3	0.0837	0.1163 (0.13)	0.1144 (0.12)	0.1264	0.1592 (0.22)	0.1549 (0.24)
	Φ_4	0.1059	0.1298 (0.32)	0.1267 (0.32)	0.0999	0.1202 (0.23)	0.1162 (0.24)
	Φ_5	0.1515	0.4070 (0.00)	0.4001 (0.01)	0.7544	0.5160 (0.11)	0.5130 (0.11)
	Φ_6	0.2030	0.4334 (0.04)	0.4273 (0.05)	0.2356	0.4526 (0.05)	0.4431 (0.05)
200	Φ_1	0.0151	0.0155 (0.46)	0.0152 (0.46)	0.0188	0.0190 (0.42)	0.0187 (0.42)
	Φ_2	0.0240	0.0246 (0.49)	0.0241 (0.48)	0.0236	0.0233 (0.57)	0.0227 (0.56)
	Φ_3	0.0397	0.0671 (0.11)	0.0659 (0.12)	0.0551	0.0975 (0.13)	0.0970 (0.14)
	Φ_4	0.0445	0.0586 (0.15)	0.0578 (0.14)	0.0562	0.0690 (0.24)	0.0677 (0.26)
	Φ_5	0.0815	0.2717 (0.01)	0.2681 (0.01)	0.1259	0.3679 (0.07)	0.3666 (0.07)
	Φ_6	0.0940	0.2163 (0.04)	0.2157 (0.04)	0.1053	0.2225 (0.07)	0.2190 (0.07)
1000	Φ_1	0.0029	0.0032 (0.36)	0.0031 (0.41)	0.0041	0.0054 (0.21)	0.0053 (0.24)
	Φ_2	0.0040	0.0043 (0.43)	0.0042 (0.46)	0.0044	0.0042 (0.56)	0.0042 (0.55)
	Φ_3	0.0082	0.0311 (0.01)	0.0309 (0.00)	0.0101	0.0432 (0.01)	0.0429 (0.01)
	Φ_4	0.0093	0.0185 (0.05)	0.0182 (0.05)	0.0108	0.0198 (0.06)	0.0197 (0.06)
	Φ_5	0.0150	0.1489 (0.00)	0.1488 (0.00)	0.0269	0.1922 (0.00)	0.1916 (0.00)
	Φ_6	0.0186	0.0816 (0.01)	0.0815 (0.01)	0.0195	0.0744 (0.01)	0.0742 (0.01)

Sample Size	Φ	Model 3 ML	Model 3 Lasso	Model 3 Lasso (OD)	Model 4 ML	Model 4 Lasso	Model 4 Lasso (OD)
50	Φ_1	0.2467	0.1692 (0.70)	0.1699 (0.69)	0.0886	0.0748 (0.69)	0.0747 (0.68)
	Φ_2	0.1471	0.1303 (0.78)	0.1303 (0.78)	0.1085	0.0994 (0.75)	0.0955 (0.81)
	Φ_3	0.7190	0.5847 (0.49)	0.5830 (0.46)	0.1976	0.2406 (0.16)	0.2366 (0.20)
	Φ_4	0.3561	0.3494 (0.39)	0.3431 (0.45)	0.2574	0.2882 (0.29)	0.2743 (0.29)
	Φ_5	5.6391	2.2247 (0.43)	2.1784 (0.41)	0.3932	0.7592 (0.08)	0.7524 (0.08)
	Φ_6	1.2695	1.2896 (0.23)	1.3148 (0.24)	0.5231	0.8445 (0.10)	0.8183 (0.10)
100	Φ_1	0.1149	0.0815 (0.70)	0.0789 (0.73)	0.0417	0.0376 (0.64)	0.0361 (0.68)
	Φ_2	0.0650	0.0562 (0.76)	0.0552 (0.75)	0.0543	0.0467 (0.73)	0.0451 (0.75)
	Φ_3	0.3440	0.3056 (0.40)	0.3054 (0.42)	0.1008	0.1221 (0.25)	0.1193 (0.24)
	Φ_4	0.1647	0.1528 (0.47)	0.1525 (0.48)	0.1112	0.1166 (0.42)	0.1153 (0.43)
	Φ_5	12.1045	1.1899 (0.28)	1.1715 (0.29)	0.1579	0.3982 (0.00)	0.3961 (0.01)
	Φ_6	0.3388	0.4764 (0.22)	0.4574 (0.24)	0.2102	0.3928 (0.04)	0.3866 (0.07)
200	Φ_1	0.0476	0.0327 (0.74)	0.0323 (0.77)	0.0219	0.0193 (0.67)	0.0190 (0.68)
	Φ_2	0.0295	0.0247 (0.83)	0.0242 (0.79)	0.0272	0.0236 (0.67)	0.0235 (0.71)
	Φ_3	0.1433	0.1999 (0.25)	0.1962 (0.27)	0.0471	0.0654 (0.18)	0.0646 (0.18)
	Φ_4	0.0731	0.0758 (0.37)	0.0738 (0.41)	0.0574	0.0656 (0.32)	0.0639 (0.33)
	Φ_5	0.5163	0.8795 (0.21)	0.8791 (0.23)	0.0757	0.2583 (0.01)	0.2558 (0.01)
	Φ_6	0.1449	0.2890 (0.09)	0.2861 (0.09)	0.1036	0.2202 (0.06)	0.2158 (0.03)
1000	Φ_1	0.0095	0.0087 (0.52)	0.0086 (0.52)	0.0037	0.0034 (0.51)	0.0033 (0.49)
	Φ_2	0.0060	0.0046 (0.74)	0.0046 (0.74)	0.0052	0.0042 (0.72)	0.0042 (0.71)
	Φ_3	0.0275	0.0754 (0.06)	0.0761 (0.07)	0.0085	0.0269 (0.00)	0.0267 (0.00)
	Φ_4	0.0148	0.0197 (0.21)	0.0197 (0.21)	0.0110	0.0180 (0.14)	0.0179 (0.14)
	Φ_5	0.0604	0.3814 (0.00)	0.3824 (0.00)	0.0155	0.1523 (0.00)	0.1522 (0.00)
	Φ_6	0.0277	0.0786 (0.01)	0.0780 (0.01)	0.0194	0.0809 (0.00)	0.0802 (0.00)

Table 6.5: Standardized Mean Squared Error Table - Lasso FA, No Model Error

sMSE : Mean Squared Error of Standardized Parameter Estimates

Sample Size	Φ	Model 1 ML	Model 1 Lasso	Model 1 Lasso (OD)	Model 2 ML	Model 2 Lasso	Model 2 Lasso (OD)
50	Φ_1	0.0692	0.0794 (0.44)	0.0760 (0.44)	0.0954	0.0955 (0.52)	0.0918 (0.56)
	Φ_2	0.0809	0.0794 (0.57)	0.0772 (0.62)	0.0988	0.1035 (0.57)	0.1020 (0.59)
	Φ_3	0.1834	0.2297 (0.19)	0.2289 (0.18)	0.2790	0.3267 (0.19)	0.3228 (0.20)
	Φ_4	0.2395	0.2694 (0.36)	0.2590 (0.37)	0.2628	0.2737 (0.35)	0.2679 (0.36)
	Φ_5	0.3253	0.7372 (0.05)	0.7204 (0.06)	1.5626	0.9835 (0.24)	0.9849 (0.24)
	Φ_6	0.4491	0.7437 (0.09)	0.7268 (0.08)	1.0555	1.0626 (0.16)	1.0490 (0.18)
100	Φ_1	0.0337	0.0352 (0.51)	0.0335 (0.57)	0.0471	0.0494 (0.34)	0.0486 (0.35)
	Φ_2	0.0407	0.0391 (0.69)	0.0384 (0.69)	0.0492	0.0461 (0.68)	0.0461 (0.70)
	Φ_3	0.0824	0.1228 (0.09)	0.1195 (0.09)	0.1409	0.2013 (0.11)	0.1969 (0.11)
	Φ_4	0.1181	0.1393 (0.24)	0.1376 (0.24)	0.1179	0.1387 (0.28)	0.1363 (0.25)
	Φ_5	0.1695	0.4556 (0.03)	0.4530 (0.04)	0.9917	0.7329 (0.08)	0.7190 (0.09)
	Φ_6	0.1930	0.4233 (0.05)	0.3965 (0.04)	0.2552	0.4921 (0.05)	0.4763 (0.06)
200	Φ_1	0.0150	0.0156 (0.52)	0.0144 (0.54)	0.0275	0.0306 (0.33)	0.0305 (0.32)
	Φ_2	0.0229	0.0223 (0.53)	0.0217 (0.54)	0.0250	0.0232 (0.68)	0.0227 (0.71)
	Φ_3	0.0384	0.0732 (0.09)	0.0718 (0.09)	0.0710	0.1414 (0.12)	0.1405 (0.12)
	Φ_4	0.0550	0.0670 (0.29)	0.0657 (0.28)	0.0618	0.0846 (0.13)	0.0832 (0.14)
	Φ_5	0.0736	0.2968 (0.02)	0.2936 (0.02)	0.1439	0.6005 (0.03)	0.5920 (0.02)
	Φ_6	0.1001	0.2522 (0.04)	0.2514 (0.04)	0.1213	0.3233 (0.03)	0.3131 (0.03)
1000	Φ_1	0.0029	0.0030 (0.40)	0.0029 (0.46)	0.0104	0.0170 (0.01)	0.0169 (0.01)
	Φ_2	0.0039	0.0037 (0.52)	0.0036 (0.57)	0.0074	0.0067 (0.71)	0.0067 (0.69)
	Φ_3	0.0084	0.0377 (0.00)	0.0375 (0.00)	0.0227	0.0927 (0.00)	0.0924 (0.00)
	Φ_4	0.0095	0.0224 (0.05)	0.0221 (0.06)	0.0152	0.0307 (0.01)	0.0306 (0.01)
	Φ_5	0.0153	0.2245 (0.00)	0.2247 (0.00)	0.0339	0.4139 (0.00)	0.4096 (0.00)
	Φ_6	0.0194	0.1148 (0.00)	0.1142 (0.00)	0.0269	0.1410 (0.00)	0.1409 (0.00)

Sample Size	Φ	Model 3 ML	Model 3 Lasso	Model 3 Lasso (OD)	Model 4 ML	Model 4 Lasso	Model 4 Lasso (OD)
50	Φ_1	0.2580	0.1955 (0.59)	0.1949 (0.63)	0.0822	0.0797 (0.65)	0.0787 (0.69)
	Φ_2	0.1662	0.1533 (0.77)	0.1531 (0.74)	0.1101	0.1022 (0.78)	0.1004 (0.77)
	Φ_3	0.7628	0.7562 (0.27)	0.7444 (0.27)	0.1985	0.2326 (0.33)	0.2323 (0.31)
	Φ_4	0.3898	0.4203 (0.36)	0.4127 (0.34)	0.2379	0.2828 (0.20)	0.2704 (0.21)
	Φ_5	81.7991	5.8647 (0.28)	5.8510 (0.29)	0.3881	0.8038 (0.06)	0.7979 (0.06)
	Φ_6	1.1118	1.1615 (0.21)	1.1443 (0.21)	0.5869	0.8204 (0.12)	0.8071 (0.13)
100	Φ_1	0.1221	0.1043 (0.63)	0.1064 (0.64)	0.0471	0.0415 (0.72)	0.0408 (0.75)
	Φ_2	0.0723	0.0639 (0.78)	0.0636 (0.82)	0.0528	0.0483 (0.78)	0.0473 (0.75)
	Φ_3	0.3520	0.4899 (0.16)	0.4861 (0.15)	0.1031	0.1308 (0.28)	0.1303 (0.28)
	Φ_4	0.2201	0.2415 (0.33)	0.2387 (0.33)	0.1185	0.1287 (0.41)	0.1280 (0.40)
	Φ_5	1.2674	1.3339 (0.13)	1.3469 (0.13)	0.1778	0.4505 (0.04)	0.4485 (0.04)
	Φ_6	0.4517	0.6882 (0.10)	0.6785 (0.09)	0.2536	0.4589 (0.14)	0.4547 (0.13)
200	Φ_1	0.0763	0.0758 (0.39)	0.0752 (0.44)	0.0212	0.0178 (0.64)	0.0177 (0.65)
	Φ_2	0.0457	0.0393 (0.88)	0.0390 (0.88)	0.0268	0.0233 (0.72)	0.0228 (0.78)
	Φ_3	0.2440	0.4305 (0.02)	0.4311 (0.02)	0.0470	0.0701 (0.19)	0.0690 (0.21)
	Φ_4	0.1298	0.1464 (0.38)	0.1444 (0.38)	0.0592	0.0698 (0.30)	0.0688 (0.28)
	Φ_5	0.6152	1.2265 (0.01)	1.2258 (0.01)	0.0817	0.3412 (0.00)	0.3379 (0.00)
	Φ_6	0.2470	0.4580 (0.15)	0.4579 (0.13)	0.0996	0.2301 (0.05)	0.2299 (0.08)
1000	Φ_1	0.0329	0.0418 (0.09)	0.0415 (0.10)	0.0039	0.0031 (0.62)	0.0031 (0.62)
	Φ_2	0.0163	0.0117 (0.90)	0.0115 (0.95)	0.0045	0.0038 (0.66)	0.0038 (0.67)
	Φ_3	0.1548	0.3067 (0.00)	0.3063 (0.00)	0.0098	0.0367 (0.00)	0.0363 (0.00)
	Φ_4	0.0721	0.0727 (0.49)	0.0723 (0.50)	0.0107	0.0217 (0.09)	0.0216 (0.07)
	Φ_5	0.2853	1.0346 (0.00)	1.0298 (0.00)	0.0157	0.2226 (0.00)	0.2215 (0.00)
	Φ_6	0.1670	0.2808 (0.04)	0.2782 (0.04)	0.0192	0.1108 (0.00)	0.1111 (0.00)

Table 6.6: Standardized Mean Squared Error Table - Lasso FA, Model Error Involved

ent method of comparison, OD estimates difference and outperformance rate, show similar trends.

Comparison on sMSE are provided in table 6.5 and 6.6. These tables suggest different tendencies with those in OD comparison. With no or less degree of misspecification, Lasso yields worse sMSE than ML's in almost all the conditions. However, as the misspecification gets severe, Lasso performs better. This trend is remarkable in that it differ considerably from the OD's tendency. Note that, in model 3 and 4, Lasso produces consistently better results for Φ_1 and Φ_2 with or without model error. For these conditions, the magnitudes of difference in outperformance rate are also not changed markedly even the model error is involved. ML generally performs better than Lasso in Φ_5 and Φ_6. However, when the sample size is not large enough, ML sometimes yields unstable results in Φ_5 condition. What should be noted is that in the same sample which shows poor sMSE, OD is not that bad. This seems due to the local minima cases, and ML suffers severely from this problem.

In sum, with regard to OD and sMSE, Comparison of Lasso and ML's performance yields different results depending on simulation conditions, especially the Φ matrices. For Φ_1 and Φ_2, Lasso outperforms ML on reducing both OD and sMSE whether the model error is involved or not. For Φ_5 and Φ_6 ML produce overwhelming results over all the conditions. For the remaining conditions, Φ_3 and Φ_4, OD result varies severely depending on the sample sizes, model error, and degree of misspecification. But sMSE is in favor of ML over almost all the conditions.

The most noteworthy point is that OD and sMSE show slightly different reaction to misspecification; OD seems not to be affected largely by misspecification whereas sMSE does.

Analyses of DA, DE, bias squared and variance are presented in Subsection

6.3.1 and Appendix A. We summarize some important results in those chapters. One thing should be noted here is that Lasso FA shows great capability of reducing DE. And this result is not affected largely by model error. However, it also turns out that the proposed method suffers from considerably larger DA than ML. It seems that excessive DA is the main factor that deteriorates Lasso's performance with regard to reducing OD.

The next thing we investigate is the estimation results of each parameters. Since the simulation studies deal with a variety of conditions and models, we cannot post all the results. Instead, mean values of 100 estimates for model 3 with $\Phi = \Phi_4$ are provided in table 6.7 and 6.8, each for without and with model error case. Furthermore, densities of those estimates in $n = 100$ condition are obtained and represented in figure 6.5 and 6.6. In those tables and plots, misspecified parameters are marked with asterisks. We select Φ_4 to inspect Lasso in a neutral position. Note that when a covariance matrix that researchers are interested in is close to the implied covariance matrix derived by Φ_1 or Φ_2 conditions, the analysis result will be more favorable to Lasso. Or, when the matrix can be assumed to generated by the process close to Φ_5 or Φ_6 conditions, the opposite will be the case.

For the case when the model error is not involved, ML and Lasso do not produce considerably different results for correctly specified parameters; ML yields better for some coefficients but Lasso also give better for the other. The noteworthy distinction occurs for the misspecified parameters. At first , Lasso yields less standard deviation than ML for misspecified parameters(check the rows highlighted in bold). This result is consistent over all the sample size conditions. Even Lasso cannot produce complete shrinkages to zero in some cases, which yields non-zero mean estimates for parameters with asterisks in the table. However, Lasso estimates' distributions in figure 6.5 are extremely dense

		Sample Size = 50		*Sample Size = 100*		*Sample Size = 200*		*Sample Size = 1,000*	
	True	ML	LASSO	ML	LASSO	ML	LASSO	ML	LASSO
λ_{21}	*0.92*	0.9322	0.8854	0.9089	0.8724	0.9195	0.8863	0.9172	0.8899
		(0.2857)	(0.2698)	(0.1398)	(0.1466)	(0.1051)	(0.1030)	(0.0473)	(0.0506)
λ_{31}	*0.90*	0.8989	0.8555	0.9046	0.8544	0.9217	0.8911	0.9040	0.8696
		(0.2344)	(0.2495)	(0.1675)	(0.1823)	(0.1073)	(0.1178)	(0.0430)	(0.0485)
λ_{41}	*1.05*	1.0465	1.0765	1.0626	1.0941	1.0625	1.0927	1.0558	1.0825
		(0.2280)	(0.3054)	(0.1408)	(0.1827)	(0.0900)	(0.1166)	(0.0485)	(0.0605)
λ_{62}	*0.98*	1.0309	1.0070	0.9914	0.9702	0.9940	0.9835	0.9779	0.9691
		(0.2363)	(0.2288)	(0.1765)	(0.1732)	(0.0976)	(0.1049)	(0.0403)	(0.0434)
λ_{72}	*1.10*	1.1404	1.1991	1.1201	1.1843	1.1108	1.1567	1.1001	1.1454
		(0.1975)	(0.2518)	(0.1543)	(0.2035)	(0.0948)	(0.1257)	(0.0410)	(0.0532)
λ_{82}	*0.90*	0.9333	0.8924	0.9289	0.8937	0.9066	0.8700	0.9041	0.8705
		(0.2444)	(0.2721)	(0.1586)	(0.1475)	(0.1124)	(0.1146)	(0.0443)	(0.0463)
λ_{61}*	-	**0.0150**	**0.0593**	**-0.0003**	**0.0426**	**-0.0057**	**0.0183**	**0.0006**	**0.0062**
		(0.1936)	**(0.1442)**	**(0.1489)**	**(0.0952)**	**(0.0740)**	**(0.0447)**	**(0.0367)**	**(0.0137)**
λ_{81}*	-	**0.0050**	**0.0553**	**0.0111**	**0.0422**	**-0.0069**	**0.0217**	**-0.0073**	**0.0028**
		(0.2136)	**(0.1463)**	**(0.1318)**	**(0.0847)**	**(0.0861)**	**(0.0414)**	**(0.0433)**	**(0.0112)**
λ_{22}*	-	**-0.0145**	**0.0378**	**0.0135**	**0.0316**	**-0.0041**	**0.0139**	**-0.0012**	**0.0028**
		(0.2262)	**(0.1525)**	**(0.1273)**	**(0.0778)**	**(0.0861)**	**(0.0453)**	**(0.0413)**	**(0.0079)**
λ_{32}*	-	**0.0295**	**0.0676**	**-0.0181**	**0.0192**	**0.0025**	**0.0182**	**-0.0029**	**0.0009**
		(0.2037)	**(0.1554)**	**(0.1292)**	**(0.0668)**	**(0.0812)**	**(0.0432)**	**(0.0358)**	**(0.0038)**
$\psi_{\epsilon 1}$	*0.40*	0.3626	0.3563	0.3847	0.3810	0.3979	0.3944	0.3987	0.3938
		(0.1192)	(0.1368)	(0.0689)	(0.0781)	(0.0568)	(0.0614)	(0.0286)	(0.0313)
$\psi_{\epsilon 2}$	*0.49*	0.4514	0.4675	0.4831	0.4919	0.4859	0.4915	0.4872	0.4898
		(0.1213)	(0.1303)	(0.0835)	(0.0897)	(0.0613)	(0.0652)	(0.0277)	(0.0287)
$\psi_{\epsilon 3}$	*0.51*	0.4828	0.4956	0.5020	0.5131	0.4991	0.5038	0.5168	0.5206
		(0.1147)	(0.1274)	(0.0851)	(0.0904)	(0.0677)	(0.0709)	(0.0312)	(0.0324)
$\psi_{\epsilon 4}$	*0.34*	0.3347	0.3260	0.3177	0.2996	0.3367	0.3178	0.3363	0.3183
		(0.1337)	(0.1504)	(0.0790)	(0.0891)	(0.0532)	(0.0591)	(0.0248)	(0.0274)
$\psi_{\epsilon 5}$	*0.40*	0.3970	0.3994	0.3939	0.3981	0.3959	0.3968	0.3978	0.3981
		(0.1206)	(0.1251)	(0.0751)	(0.0813)	(0.0578)	(0.0628)	(0.0224)	(0.0237)
$\psi_{\epsilon 6}$	*0.42*	0.3863	0.3955	0.4109	0.4195	0.4065	0.4097	0.4199	0.4222
		(0.1089)	(0.1182)	(0.0928)	(0.0949)	(0.0572)	(0.0614)	(0.0254)	(0.0266)
$\psi_{\epsilon 7}$	*0.27*	0.2523	0.2341	0.2685	0.2469	0.2801	0.2611	0.2761	0.2570
		(0.1001)	(0.1057)	(0.0695)	(0.0786)	(0.0486)	(0.0557)	(0.0214)	(0.0233)
$\psi_{\epsilon 8}$	*0.51*	0.4760	0.4909	0.4795	0.4898	0.5026	0.5107	0.5113	0.5173
		(0.1157)	(0.1254)	(0.0881)	(0.0939)	(0.0634)	(0.0649)	(0.0276)	(0.0281)
ϕ_{11}	*0.60*	0.6113	0.6253	0.6254	0.6389	0.5852	0.6113	0.5973	0.6257
		(0.2220)	(0.2390)	(0.1506)	(0.1452)	(0.0885)	(0.0873)	(0.0482)	(0.0485)
ϕ_{12}	*0.45*	0.2144	0.1731	0.2294	0.1904	0.2169	0.1872	0.2229	0.2012
		(0.1051)	(0.0844)	(0.0814)	(0.0680)	(0.0534)	(0.0409)	(0.0265)	(0.0220)
ϕ_{22}	*0.60*	0.5764	0.5629	0.5788	0.5678	0.5904	0.5977	0.6009	0.6083
		(0.1755)	(0.1845)	(0.1350)	(0.1379)	(0.0883)	(0.0897)	(0.0429)	(0.0415)

Table 6.7: Estimates of Parameters - FA, Model 3, No Model Error, Φ_4

around zero. As a matter of fact, the result is due to the complete shrinkage of the nuisance parameters generated by Lasso. This will be shown in the table 6.9 and 6.10.

Before scrutinizing tables regarding the shrinkage effect, we investigate the Lasso's estimation results for the case when the model error is involved. Ta-

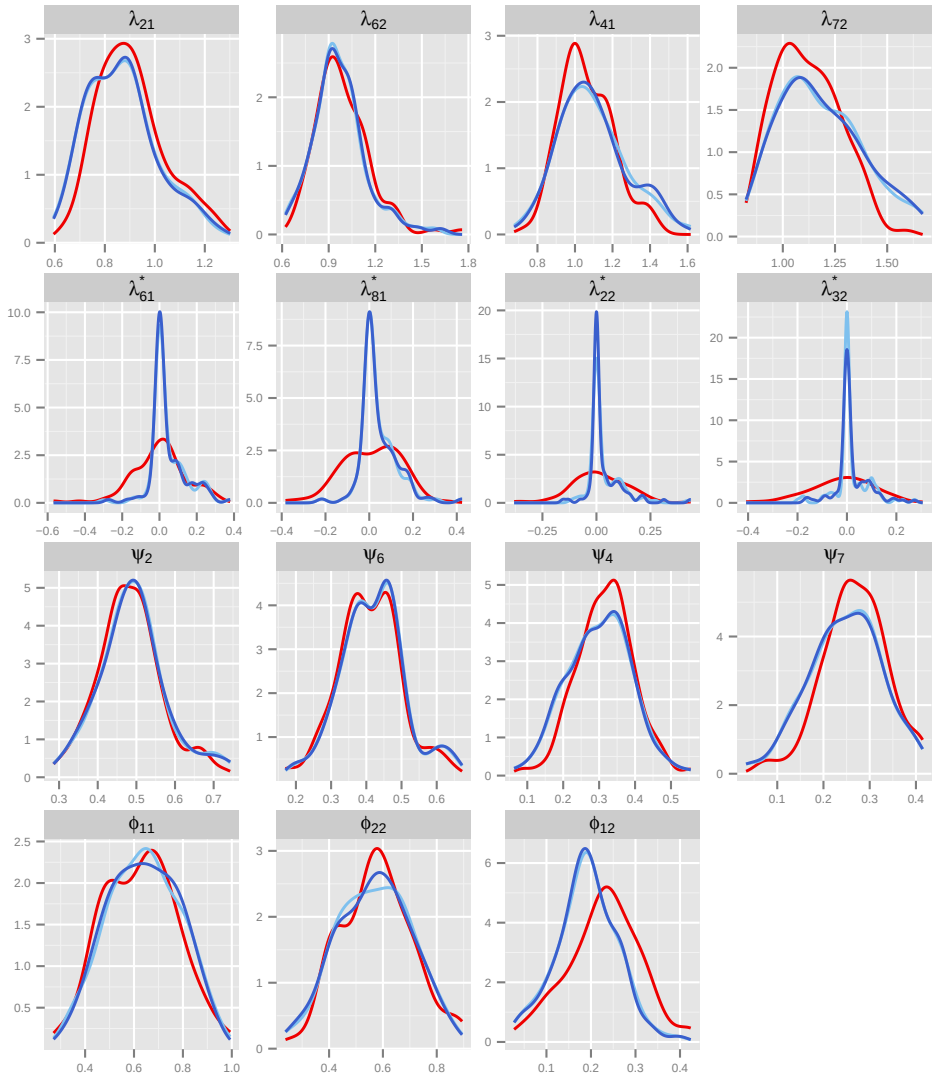

Figure 6.5: Density Plots for Selected Parameters - FA, Model 3, No Model Error, $N = 100$, Φ_4

ble 6.8 and figure 6.6 do not make positive contributions for Lasso. In the table, both ML and Lasso's estimates for misspecified parameters in crease to a non-ignorable degree. Also standard deviations get larger, especially for Lasso.

		Sample Size = 50		*Sample Size = 100*		*Sample Size = 200*		*Sample Size = 1,000*	
	True	ML	LASSO	ML	LASSO	ML	LASSO	ML	LASSO
λ_{21}	*0.92*	0.9401	0.9243	0.9112	0.8928	0.8811	0.8739	0.8899	0.8867
		(0.2339)	(0.2670)	(0.1814)	(0.1987)	(0.0971)	(0.1066)	(0.0499)	(0.0553)
λ_{31}	*0.90*	0.8881	0.8700	0.8610	0.8516	0.8561	0.8457	0.8624	0.8564
		(0.2503)	(0.2684)	(0.1708)	(0.1824)	(0.1048)	(0.1157)	(0.0444)	(0.0506)
λ_{41}	*1.05*	1.1106	1.1586	1.0861	1.1292	1.0693	1.0969	1.0705	1.0986
		(0.2585)	(0.3142)	(0.1687)	(0.2098)	(0.0981)	(0.1238)	(0.0492)	(0.0602)
λ_{62}	*0.98*	0.9474	0.9441	0.9525	0.9523	0.9266	0.9286	0.9397	0.9449
		(0.2366)	(0.2526)	(0.1357)	(0.1464)	(0.0995)	(0.1067)	(0.0464)	(0.0506)
λ_{72}	*1.10*	1.1029	1.1668	1.1278	1.1899	1.1224	1.1767	1.1204	1.1758
		(0.1833)	(0.2226)	(0.1510)	(0.1907)	(0.0908)	(0.1137)	(0.0454)	(0.0575)
λ_{82}	*0.90*	0.8410	0.8169	0.8706	0.8585	0.8281	0.8164	0.8401	0.8284
		(0.2152)	(0.2208)	(0.1255)	(0.1409)	(0.0977)	(0.1041)	(0.0470)	(0.0513)
λ_{61}*	-	**0.0796**	**0.1115**	**0.0960**	**0.1014**	**0.1133**	**0.1027**	**0.1096**	**0.0888**
		(0.1980)	**(0.1698)**	**(0.1284)**	**(0.1103)**	**(0.0808)**	**(0.0850)**	**(0.0356)**	**(0.0415)**
λ_{81}*	-	**0.1136**	**0.1193**	**0.1598**	**0.1460**	**0.1570**	**0.1337**	**0.1605**	**0.1308**
		(0.2049)	**(0.1773)**	**(0.1213)**	**(0.1102)**	**(0.0839)**	**(0.0887)**	**(0.0413)**	**(0.0420)**
λ_{22}*	-	**0.0715**	**0.0910**	**0.0751**	**0.0809**	**0.0942**	**0.0682**	**0.0977**	**0.0529**
		(0.1707)	**(0.1381)**	**(0.1453)**	**(0.1176)**	**(0.0783)**	**(0.0723)**	**(0.0374)**	**(0.0387)**
λ_{32}*	-	**0.1109**	**0.1189**	**0.1342**	**0.1225**	**0.1233**	**0.0943**	**0.1186**	**0.0708**
		(0.1953)	**(0.1746)**	**(0.1386)**	**(0.1259)**	**(0.0826)**	**(0.0805)**	**(0.0348)**	**(0.0425)**
$\psi_{\epsilon 1}$	*0.40*	0.3917	0.3918	0.3851	0.3856	0.3933	0.3917	0.4004	0.3983
		(0.1152)	(0.1280)	(0.0925)	(0.0984)	(0.0591)	(0.0661)	(0.0266)	(0.0285)
$\psi_{\epsilon 2}$	*0.49*	0.4677	0.4791	0.4826	0.4916	0.4813	0.4838	0.4973	0.4984
		(0.1138)	(0.1271)	(0.0853)	(0.0907)	(0.0563)	(0.0589)	(0.0248)	(0.0259)
$\psi_{\epsilon 3}$	*0.51*	0.4676	0.4777	0.4915	0.4965	0.5042	0.5079	0.5157	0.5175
		(0.1135)	(0.1236)	(0.0717)	(0.0764)	(0.0587)	(0.0618)	(0.0286)	(0.0298)
$\psi_{\epsilon 4}$	*0.34*	0.3122	0.2943	0.3204	0.2994	0.3190	0.3021	0.3259	0.3086
		(0.1263)	(0.1382)	(0.0813)	(0.0920)	(0.0518)	(0.0586)	(0.0233)	(0.0265)
$\psi_{\epsilon 5}$	*0.40*	0.3873	0.3986	0.3868	0.3942	0.3904	0.3967	0.4006	0.4067
		(0.0930)	(0.1029)	(0.0816)	(0.0887)	(0.0545)	(0.0583)	(0.0217)	(0.0235)
$\psi_{\epsilon 6}$	*0.42*	0.3735	0.3810	0.4067	0.4131	0.4235	0.4275	0.4269	0.4298
		(0.1144)	(0.1209)	(0.0695)	(0.0721)	(0.0518)	(0.0535)	(0.0232)	(0.0243)
$\psi_{\epsilon 7}$	*0.27*	0.2429	0.2206	0.2421	0.2182	0.2412	0.2176	0.2522	0.2276
		(0.1160)	(0.1232)	(0.0717)	(0.0775)	(0.0458)	(0.0532)	(0.0230)	(0.0273)
$\psi_{\epsilon 8}$	*0.51*	0.4648	0.4786	0.5051	0.5142	0.5035	0.5109	0.5137	0.5198
		(0.1276)	(0.1399)	(0.0759)	(0.0797)	(0.0622)	(0.0658)	(0.0276)	(0.0286)
ϕ_{11}	*0.60*	0.5893	0.5917	0.5900	0.6083	0.5962	0.6332	0.5930	0.6317
		(0.1920)	(0.2104)	(0.1436)	(0.1588)	(0.1033)	(0.1013)	(0.0465)	(0.0462)
ϕ_{12}	*0.45*	0.1808	0.1421	0.1704	0.1447	0.1704	0.1582	0.1721	0.1660
		(0.1220)	(0.0979)	(0.0918)	(0.0803)	(0.0527)	(0.0531)	(0.0242)	(0.0255)
ϕ_{22}	*0.60*	0.6208	0.5936	0.5959	0.5906	0.6063	0.6149	0.5967	0.6082
		(0.2245)	(0.2046)	(0.1494)	(0.1424)	(0.0987)	(0.0940)	(0.0434)	(0.0434)

Table 6.8: Estimates of Parameters - FA, Model 3, Model Error Involved, Φ_4

Similar tendencies can be found in the density plots. Comparing to the no model error case, distributions for misspecified parameters are concentrated around zero only in a slight degree. It seems that estimates for those parameters are observed frequently between 0 and 0.3, which produces slightly wide or bimodal(especially for λ_{81}) distributions. This result is caused by model

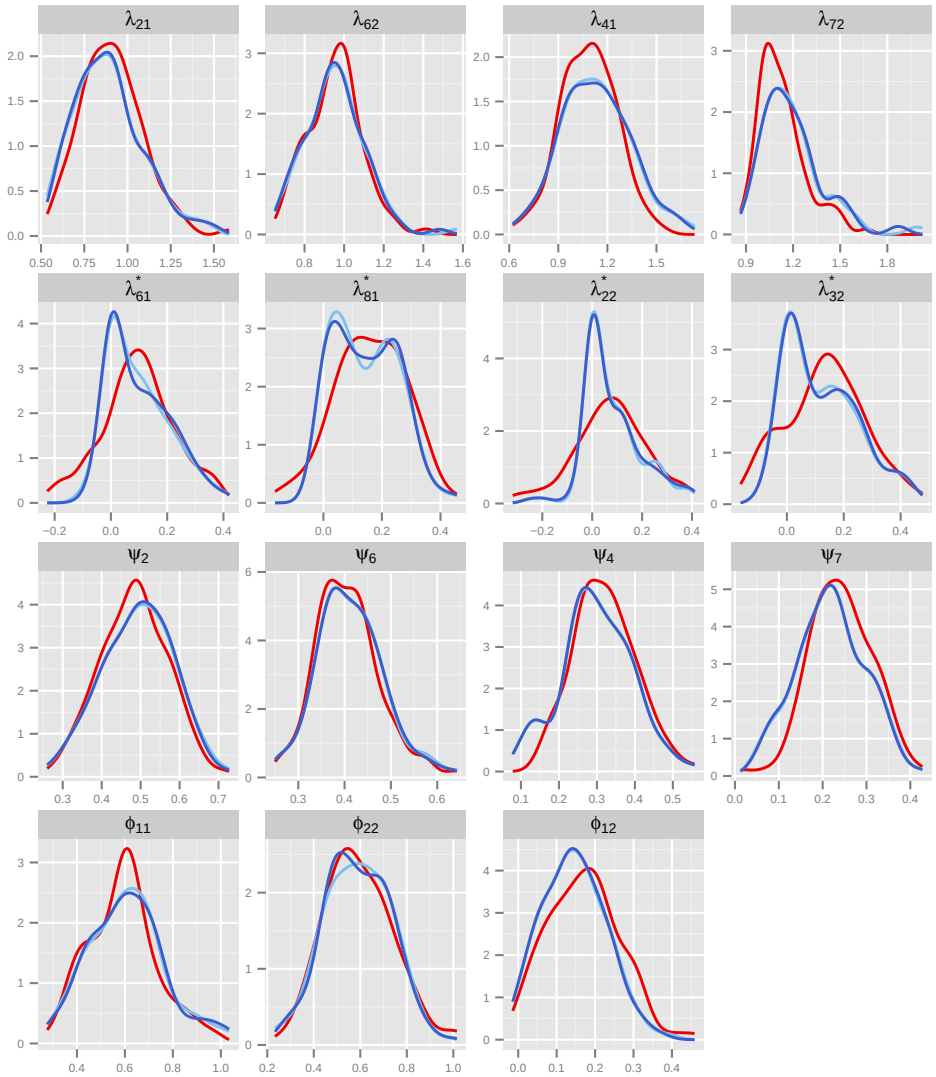

Figure 6.6: Density Plots for Selected Parameters - FA, Model 3, Model Error Involved, $N = 100$, Φ_4

error which mixed with variabilities generated by true relationship in the population. It seems that the current method to implement Lasso to FA is not good enough to distinguish meaningful covariation and model error so that it can

		Without Model Error								With Model Error							
Φ		LASSO				LASSO.OD				LASSO				LASSO.OD			
		50	*100*	*200*	*1000*	*50*	*100*	*200*	*1000*	*50*	*100*	*200*	*1000*	*50*	*100*	*200*	*1000*
Φ_1	λ_{61}	0.22	0.25	0.30	0.17	0.21	0.25	0.28	0.19	0.12	0.14	0.11	0.00	0.09	0.13	0.10	0.00
	λ_{81}	0.37	0.31	0.36	0.26	0.36	0.33	0.31	0.24	0.25	0.16	0.08	0.00	0.23	0.15	0.06	0.00
	λ_{22}	0.31	0.37	0.41	0.54	0.28	0.37	0.41	0.54	0.24	0.26	0.43	0.44	0.27	0.24	0.46	0.46
	λ_{32}	0.30	0.37	0.45	0.46	0.29	0.35	0.44	0.47	0.27	0.36	0.32	0.35	0.29	0.29	0.33	0.39
Φ_2	λ_{61}	0.21	0.41	0.36	0.54	0.19	0.41	0.35	0.55	0.16	0.18	0.17	0.02	0.16	0.18	0.18	0.03
	λ_{81}	0.35	0.42	0.38	0.74	0.32	0.43	0.40	0.76	0.25	0.18	0.13	0.05	0.24	0.19	0.13	0.05
	λ_{22}	0.32	0.41	0.40	0.70	0.27	0.44	0.41	0.69	0.30	0.28	0.33	0.39	0.34	0.29	0.31	0.44
	λ_{32}	0.21	0.47	0.40	0.80	0.20	0.44	0.39	0.80	0.29	0.28	0.27	0.36	0.28	0.31	0.28	0.39
Φ_3	λ_{61}	0.20	0.21	0.22	0.05	0.19	0.20	0.21	0.05	0.08	0.05	0.05	0.00	0.09	0.05	0.05	0.00
	λ_{81}	0.24	0.25	0.24	0.15	0.26	0.24	0.23	0.16	0.12	0.02	0.01	0.00	0.09	0.03	0.01	0.00
	λ_{22}	0.23	0.30	0.35	0.26	0.26	0.30	0.36	0.25	0.25	0.14	0.14	0.01	0.22	0.13	0.15	0.01
	λ_{32}	0.23	0.19	0.33	0.29	0.25	0.17	0.33	0.27	0.15	0.08	0.15	0.00	0.14	0.07	0.13	0.00
Φ_4	λ_{61}	0.30	0.45	0.62	0.75	0.27	0.46	0.64	0.77	0.26	0.27	0.18	0.00	0.21	0.26	0.16	0.01
	λ_{81}	0.37	0.38	0.59	0.90	0.34	0.39	0.62	0.87	0.21	0.16	0.10	0.00	0.20	0.16	0.05	0.00
	λ_{22}	0.36	0.53	0.65	0.83	0.31	0.52	0.68	0.82	0.27	0.36	0.25	0.11	0.25	0.36	0.21	0.12
	λ_{32}	0.35	0.53	0.63	0.91	0.34	0.53	0.61	0.90	0.26	0.25	0.23	0.06	0.23	0.25	0.18	0.06
Φ_5	λ_{61}	0.21	0.27	0.19	0.12	0.21	0.27	0.20	0.13	0.19	0.15	0.11	0.00	0.16	0.17	0.12	0.00
	λ_{81}	0.21	0.17	0.27	0.11	0.20	0.21	0.26	0.12	0.11	0.05	0.13	0.00	0.13	0.07	0.12	0.00
	λ_{22}	0.19	0.26	0.24	0.08	0.20	0.27	0.25	0.09	0.16	0.19	0.22	0.24	0.16	0.20	0.20	0.23
	λ_{32}	0.20	0.24	0.29	0.17	0.19	0.23	0.32	0.18	0.12	0.24	0.14	0.09	0.09	0.25	0.14	0.09
Φ_6	λ_{61}	0.42	0.49	0.57	0.95	0.40	0.49	0.57	0.92	0.27	0.28	0.26	0.07	0.26	0.24	0.29	0.08
	λ_{81}	0.37	0.43	0.60	0.95	0.38	0.43	0.62	0.97	0.24	0.19	0.09	0.00	0.21	0.16	0.11	0.00
	λ_{22}	0.31	0.56	0.59	0.96	0.32	0.53	0.60	0.96	0.23	0.33	0.26	0.15	0.22	0.33	0.27	0.14
	λ_{32}	0.38	0.51	0.63	0.96	0.38	0.50	0.65	0.97	0.23	0.22	0.22	0.07	0.23	0.22	0.22	0.05

Table 6.9: Proportion of Complete Shrinkage to Zero - Lasso FA, Model 3

detect and remove misspecified parameters.

This interpretation, regarding the effect of model error on the shrinkage effect of Lasso, is corroborated by the following tables. Table 6.9 contains the pro-

portion of Lasso's complete shrinkage observed in model 3. Cells are colored with different color intensity according to the size of occurrence rate. The table shows remarkable difference between no model error case and model error-involved case. When the model error is not contained in data-generating process, misspecified coefficients are detected and removed fairly well by Lasso. However, in case the model error is considered, the result is reversed dramatically. The proportion of complete shrinkage to zero drops by about a third.

Other points are related to the fact that the shrinkage results vary depending on sample sizes and various Φ matrices. At first, Φ_2, Φ_4, and Φ_6 yield bet-

Misspecified λ_{61} *in Model 2*

		Without Model Error								With Model Error							
Φ		LASSO				LASSO.OD				LASSO				LASSO.OD			
		50	*100*	*200*	*1000*	*50*	*100*	*200*	*1000*	*50*	*100*	*200*	*1000*	*50*	*100*	*200*	*1000*
Φ_1	λ_{61}	0.29	0.24	0.29	0.21	0.25	0.21	0.27	0.17	0.19	0.10	0.05	0.00	0.17	0.08	0.07	0.00
Φ_2		0.21	0.27	0.28	0.69	0.19	0.23	0.27	0.67	0.23	0.19	0.11	0.03	0.23	0.18	0.10	0.03
Φ_3		0.19	0.25	0.10	0.05	0.17	0.25	0.12	0.05	0.11	0.12	0.03	0.00	0.10	0.13	0.04	0.00
Φ_4		0.30	0.48	0.59	0.82	0.29	0.49	0.60	0.82	0.23	0.22	0.26	0.07	0.19	0.22	0.24	0.05
Φ_5		0.14	0.27	0.19	0.11	0.14	0.26	0.20	0.13	0.12	0.07	0.05	0.00	0.12	0.08	0.03	0.00
Φ_6		0.32	0.52	0.68	0.95	0.34	0.49	0.69	0.95	0.32	0.32	0.32	0.25	0.28	0.27	0.32	0.26

Misspecified λ_{92} *in Model 4*

		Without Model Error								With Model Error							
Φ		LASSO				LASSO.OD				LASSO				LASSO.OD			
		50	*100*	*200*	*1000*	*50*	*100*	*200*	*1000*	*50*	*100*	*200*	*1000*	*50*	*100*	*200*	*1000*
Φ_1	λ_{92}	0.43	0.46	0.47	0.76	0.40	0.46	0.46	0.79	0.37	0.39	0.43	0.65	0.39	0.37	0.45	0.65
Φ_2		0.38	0.44	0.62	0.76	0.33	0.42	0.59	0.79	0.43	0.37	0.48	0.67	0.42	0.42	0.46	0.66
Φ_3		0.43	0.47	0.55	0.93	0.40	0.44	0.55	0.92	0.30	0.45	0.53	0.82	0.29	0.46	0.53	0.84
Φ_4		0.27	0.54	0.71	0.99	0.25	0.52	0.70	0.99	0.40	0.56	0.59	0.98	0.40	0.56	0.64	0.98
Φ_5		0.29	0.48	0.77	1.00	0.32	0.49	0.75	1.00	0.29	0.41	0.65	0.92	0.28	0.39	0.64	0.94
Φ_6		0.36	0.57	0.87	1.00	0.36	0.55	0.86	1.00	0.34	0.57	0.85	1.00	0.35	0.58	0.86	1.00

Table 6.10: Proportion of Complete Shrinkage to Zero - Lasso FA, Model 2, 4

ter results than Φ_1, Φ_3, and Φ_5. Lasso performs worst in Φ_3 and Φ_5. This is reasonable since the former contains equal variances but smaller covariances with those in the latter conditions. Thus, variables belong to different latent variables have weak covariational relations in the former conditions. This enables Lasso to detect the misspecified parameters and shrink their coefficients to zero.

Secondly, the effect of larger sample sizes on Lasso's shrinkage property differs depending on whether the model error is involved or not. When the model error is not assumed, shrinkage occurs more correctly as the sample size is increasing. However, with model error, the opposite results are obtained especially for $\Phi_3 \sim \Phi_6$; occurrence rates of complete shrinkage tend to decrease as n gets larger.

Table 6.10 displays the same material as the previous one, but for model 2 and 4. In model 2, the same trend can be observed; Lasso demonstrates its ability when the model error is not contained in the covariance, but this property becomes less observable in model error case. The Varying effect of sample sizes also appears. However, shrinkage rate in model 4 shows a different result. Here Lasso discovers unnecessary parameters well whether we assume the model error or not. Note that in model 4, we engage a new variable to the model which is completely irrelevant in the population. Since this variable is not affected by the model error, the result is not surprising. Another point attracting attention is that the shrinkage trend is quite even among all the conditions, even for Φ_3 and Φ_5. This is reasonable since the ϕ matrix defines relationship among the original variables, not a newly one.

In analysis of BLasso's shrinkage effect, we tested a strategy that judges parameters to be shrunken completely to zero if they are less than a pre-determined bound. The same strategy can also be applied to Lasso, and the results are

posted in table 6.11 and 6.12. The results are far better than the other two methods - ML and BLasso. The most noticeable point is that Lasso is able to produce complete shrinkage even when it is judged with quite small bounds. Note that the numbers of complete shrinkage do not change when the bound is less than 10^{-3}. In fact, these numbers are almost equal to those we obtained using 10^{-15} as a bound; the difference is trivial, if any. Therefore, they can be accepted to indicate the number of cases that the misspecified parameters become exactly zero.

However, the numbers of occurrence of complete shrinkage decline by less than a half as the model error is added. Again, this confirms the negative effect of model error on complete shrinkage effect of Lasso.

The result of study on Lasso's shrinkage effect might seem dismal, vitiating the meaning and significance of implementing L_1-regularization to FA and SEM. However, it should be noted that the lack of shrinkage ability can be remedied by exploiting different optimization criteria. More detailed discussion on this point will be proposed in the next chapter.

Shrinakge Effect in Lasso FA : Without Model Error

	Model 2				*Model 3*				*Model 4*			
	Total Number of Trials		2,400		Total Number of Trials		9,600		Total Number of Trials		2,400	
	Number of Non-convergence		4		Number of Non-convergence		112		Number of Non-convergence		-	
	Number of Shrinkage Cases		2,396		Number of Shrinkage Cases		9,488		Number of Shrinkage Cases		2,400	
Bound	**ML**		**Lasso**		**ML**		**Lasso**		**ML**		**Lasso**	
10^{-1}	1,486	(62.3%)	1,678	(70.0%)	5,612	(59.0%)	6,759	(71.2%)	1,486	(61.9%)	2,030	(67.5%)
10^{-2}	230	(9.60%)	946	(39.5%)	837	(8.8%)	4,230	(44.6%)	181	(7.54%)	1,522	(63.4%)
10^{-3}	26	(1.09%)	843	(35.2%)	92	(0.97%)	3,863	(40.7%)	11	(0.46%)	1,443	(60.1%)
10^{-4}	1	(0.04%)	828	(34.6%)	10	(0.11%)	3,823	(40.3%)	1	(0.04%)	1,435	(59.8%)
10^{-5}	-	-	825	(34.4%)	-	-	3,820	(40.3%)	-	-	1,434	(59.8%)
⋮			⋮				⋮				⋮	
10^{-15}	-	-	825	(34.4%)	-	-	3,819	(40.3%)	-	-	1,434	(59.8%)

Table 6.11: Trends of Shrinkage Effect in Lasso FA - Without Model Error

Shrinakge Effect in Lasso FA : With Model Error δ

	Model 2				*Model 3*				*Model 4*			
	Total Number of Trials		2,400		Total Number of Trials		9,600		Total Number of Trials		2,400	
	Number of Non-convergence		4		Number of Non-convergence		112		Number of Non-convergence		-	
	Number of Shrinkage Cases		2,396		Number of Shrinkage Cases		9,488		Number of Shrinkage Cases		2,400	
Bound	**ML**		**Lasso**		**ML**		**Lasso**		**ML**		**Lasso**	
10^{-1}	1,211	(50.5%)	1,166	(48.6%)	3,631	(38.0%)	4,636	(48.6%)	1,526	(63.6%)	1,998	(83.3%)
10^{-2}	97	(4.04%)	385	(16.0%)	372	(3.90%)	1,886	(19.8%)	227	(9.5%)	1,448	(60.4%)
10^{-3}	13	(0.54%)	329	(13.7%)	43	(0.45%)	1,653	(17.3%)	22	(0.92%)	1,362	(56.8%)
10^{-4}	-	-	319	(13.3%)	5	(0.05%)	1,633	(17.1%)	-	-	1,357	(56.6%)
10^{-5}	-	-	319	(13.3%)	-	-	1,633	(17.1%)	-	-	1,357	(56.6%)
⋮			⋮				⋮				⋮	
10^{-15}	-	-	319	(13.3%)	-	-	1,633	(17.1%)	-	-	1,356	(56.5%)

Table 6.12: Trends of Shrinkage Effect in Lasso FA - With Model Error δ

6.2 Research 2: Structural Equation Model

This section will present the simulation result for the structural equation

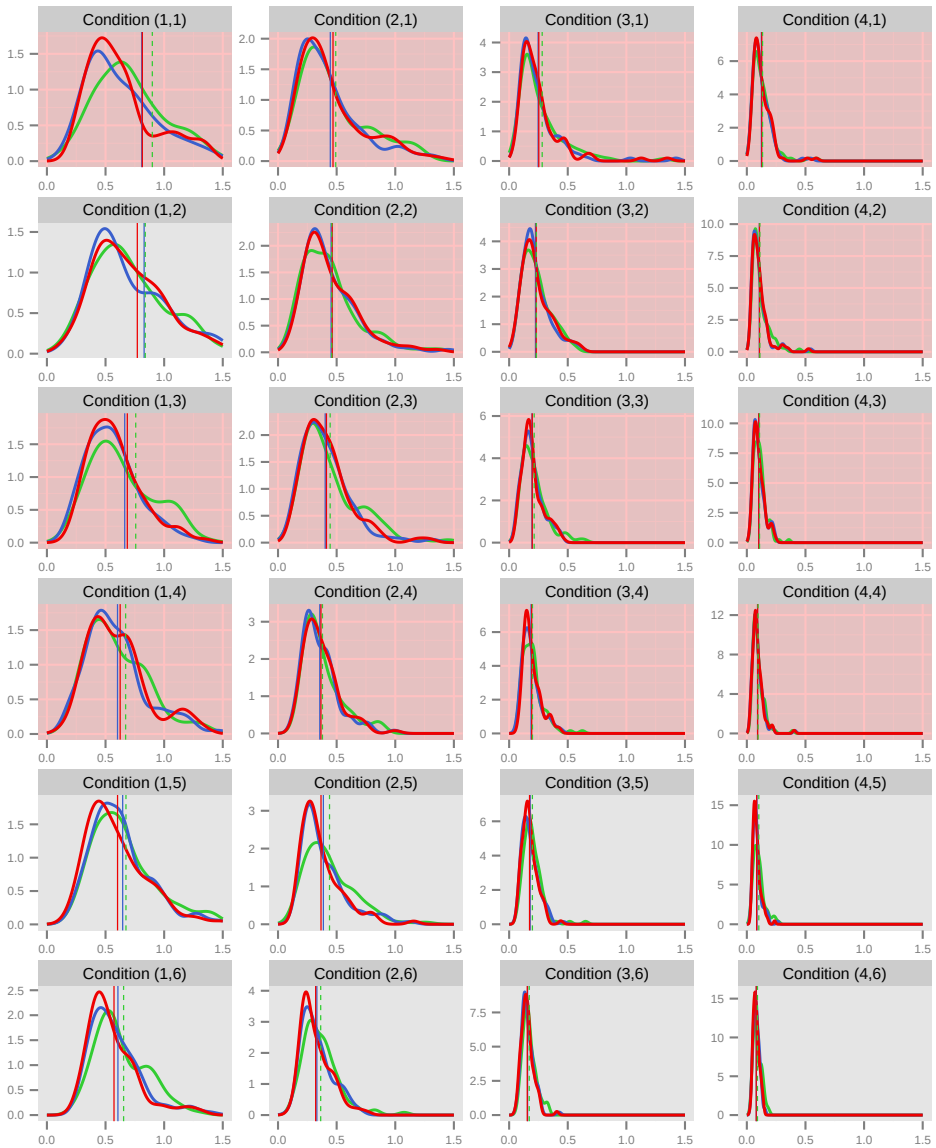

Figure 6.7: Density Plots of Overall Discrepancy: BLasso, No Model Error

models. Figure 6.7 and 6.8 contain the OD density plots for SEM model 3. An interesting point is that, in contrast to BLasso FA, posterior means outperform MAPs in BLasso SEM with respect to reducing mean OD. Thus, we use the

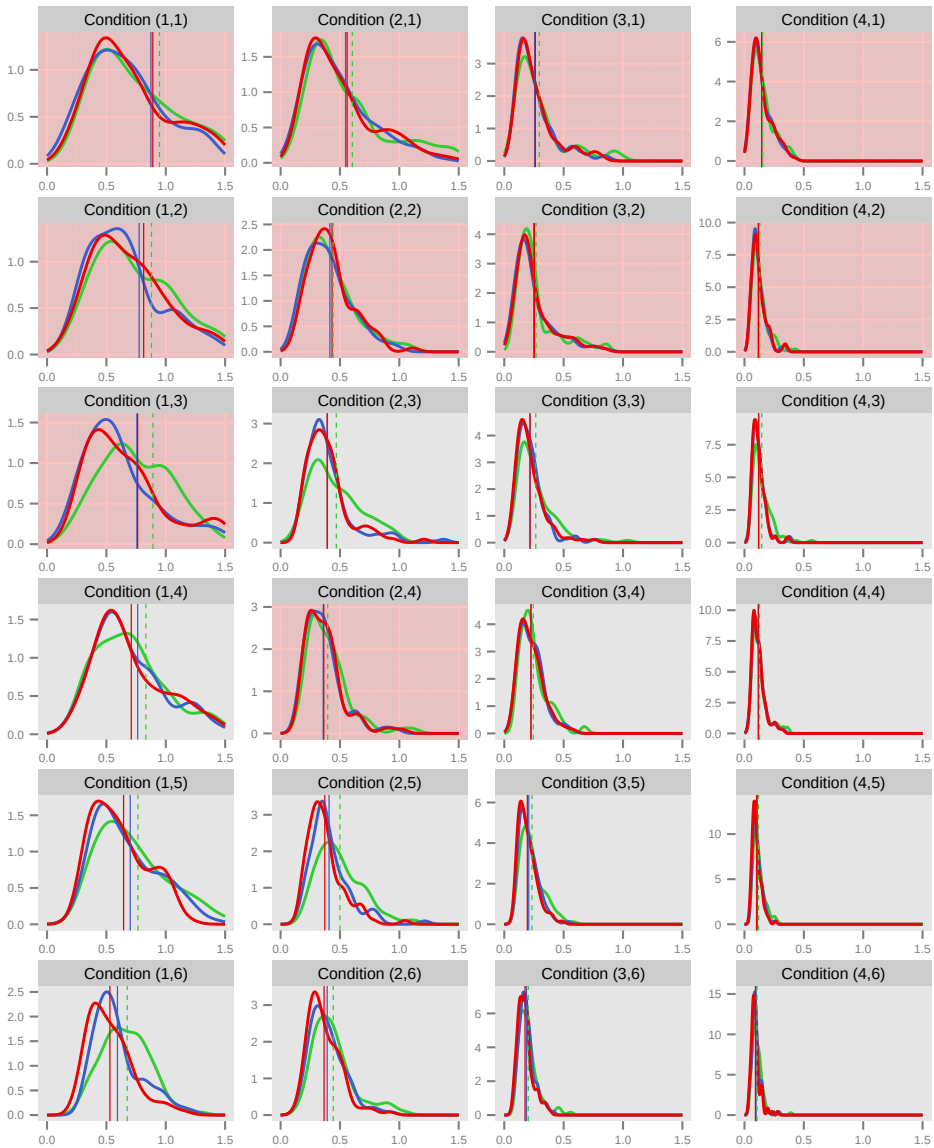

Figure 6.8: Density Plots of Overall Discrepancy: BLasso, Model Error Involved

posterior mean when comparing BLasso with ML. Note that the result from posterior mean is represented with blue color. MAP results are also posted additionally, with green color and a dotted mean line. Figures imply that BLasso performs better than OD if we do not employ the model error. However, when the model error is considered, the result changes greatly. Even though we do not post the results of other models here, the overall tendencies as follows. When the model error is not included in the analysis, BLasso consistently produce good outcomes for Φ_1 and Φ_2, except in model 2 where the method acts badly when the sample size is 150. BLasso's performance becomes better as the degree of misspecification is increased. For Φ_3 and Φ_4, the result varies so severely that any systematic interpretation is not available. For Φ_5 and Φ_6, ML always performs better. Tables containing mean estimates of OD are provided in Appendix A.

When the model error is involved, things become even worse for BLasso. It yields good results only for Φ_1 and Φ_2, and in the other Φ conditions ML always produces better results.

Table 6.13 and 6.14 contains the numbers and rates of complete shrinkages assessed by the strategy using various pre-determined bounds. As in the BLasso FA, BLasso SEM is not able to produce any outcome which is able to convince us that the strategy is useful. Most coefficient estimates are larger than 10^{-2}. Things get worse when the model error is engaged in population-generating process. Furthermore, differences between ML and BLasso become much weaker in SEM than in FA. This result rejects a need for implementing BLasso, instead of ML, to detect and remove unnecessary parameters.

For comparing the proposed Lasso SEM with ML, figure 6.9 and 6.10 provide the OD plots obtained in model 3. When the OD mean estimate is smaller in Lasso, the corresponding panels are colored red. Values of these estimates

Shrinakge Effect in BLasso SEM : Without Model Error

	Model 2				*Model 3*			
	Total Number of Trials		2,400		Total Number of Trials		9,600	
	Number of Non-convergence		1		Number of Non-convergence		12	
	Number of Shrinkage Cases		2,399		Number of Shrinkage Cases		9,588	
Bound	**ML**		**BLasso**		**ML**		**BLasso**	
10^{-1}	1,547	(64.5%)	1,690	(70.5%)	6,702	(69.9%)	7,828	(81.6%)
10^{-2}	220	(9.17%)	243	(10.1%)	915	(9.54%)	1,232	(12.9%)
10^{-3}	15	(0.63%)	22	(0.92%)	87	(0.91%)	126	(1.31%)
10^{-4}	4	(0.17%)	2	(0.08%)	11	(0.11%)	13	(0.14%)
10^{-5}	-	-	1	(0.04%)	4	(0.04%)	2	(0.02%)
10^{-6}	-	-	-	-	-	-	-	-
10^{-7}	-	-	-	-	-	-	-	-

Table 6.13: Shrinkage Effect in BLasso SEM: Without Model Error

Shrinakge Effect in BLasso SEM : With Model Error δ

	Model 2				*Model 3*			
	Total Number of Trials		2,400		Total Number of Trials		9,600	
	Number of Non-convergence		-		Number of Non-convergence		-	
	Number of Shrinkage Cases		2,400		Number of Shrinkage Cases		9,600	
Bound	**ML**		**BLasso**		**ML**		**BLasso**	
10^{-1}	1,027	(42.8%)	1,025	(42.7%)	3,866	(40.3%)	4,384	(45.7%)
10^{-2}	105	(4.38%)	150	(6.25%)	396	(4.13%)	755	(7.86%)
10^{-3}	17	(0.71%)	23	(0.96%)	41	(0.43%)	78	(0.81%)
10^{-4}	2	(0.08%)	2	(0.08%)	3	(0.03%)	7	(0.07%)
10^{-5}	-	-	-	-	-	-	-	-
10^{-6}	-	-	-	-	-	-	-	-
10^{-7}	-	-	-	-	-	-	-	-

Table 6.14: Shrinkage Effect in BLasso SEM: With Model Error δ

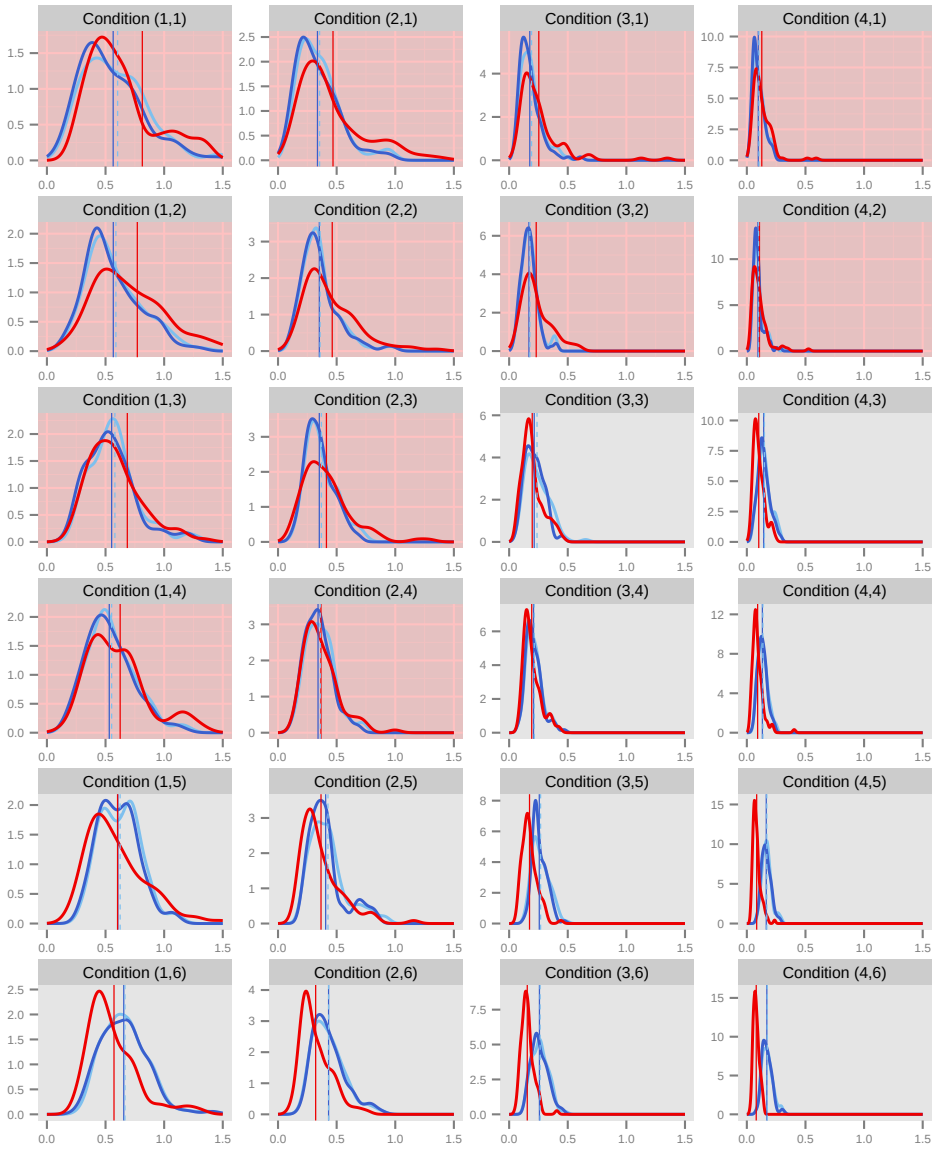

Figure 6.9: Density Plots of Overall Discrepancy : Lasso SEM, No Model Error

are also posted in table 6.15 and 6.16 for all the models and simulation conditions. Among those tables and figures, the former present the case we assume there is no model error, and the latter contains the result of model error case.

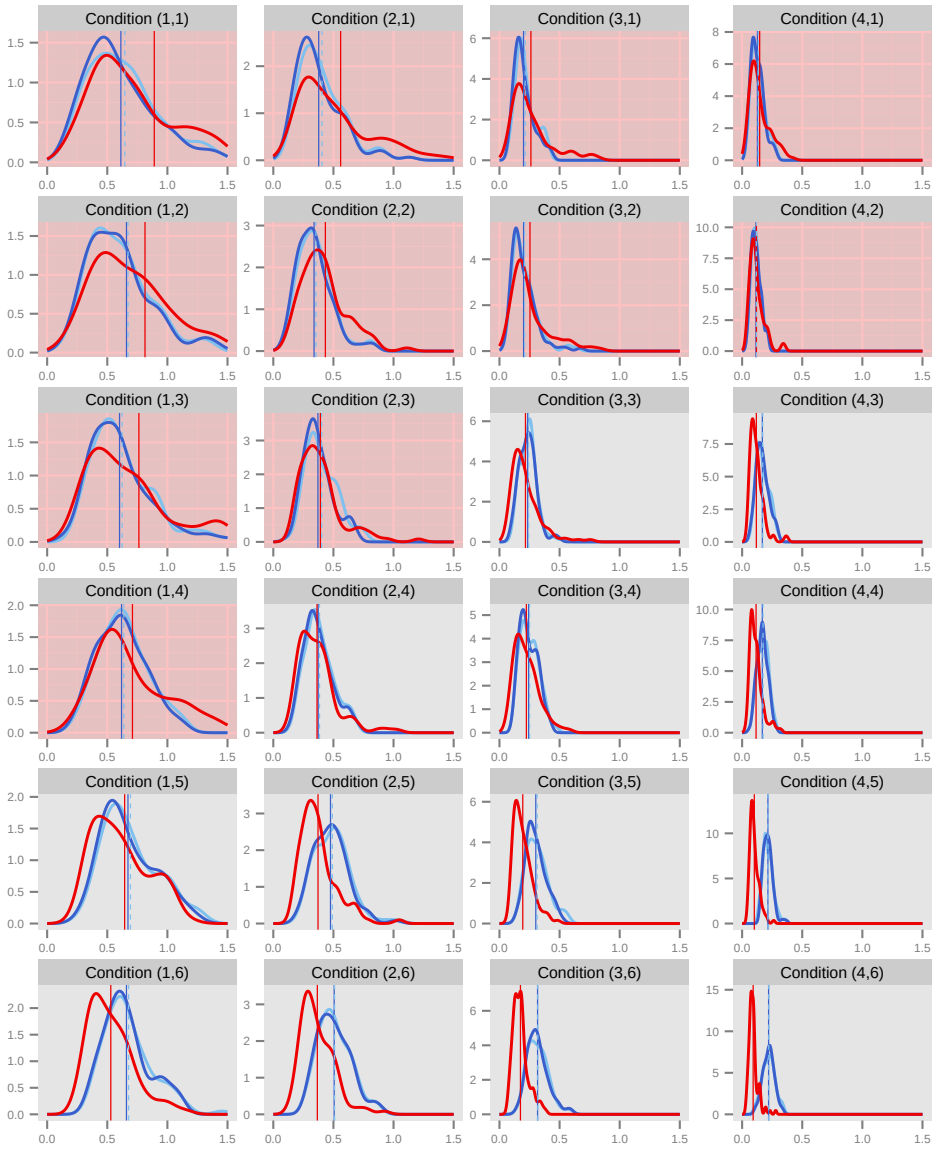

Figure 6.10: Density Plots of Overall Discrepancy : Lasso SEM, Model Error Involved

The result seems not be affected severely by the model error. In both cases, Lasso yields better outcomes than ML for Φ_1 and Φ_2 based on the values of OD estimates and outperformance rates. In contrast, ML overwhelms Lasso

OD : Overall Discrepancy (OLS)

Sample Size	Φ	Model 1 ML	Model 1 Lasso	Model 1 Lasso (OD)	Model 2 ML	Model 2 Lasso	Model 2 Lasso (OD)
50	Φ_1	0.8328	0.6190 (0.71)	0.5755 (0.82)	0.7906	0.6031 (0.72)	0.5757 (0.80)
	Φ_2	0.7107	0.5714 (0.72)	0.5473 (0.76)	0.7024	0.5236 (0.79)	0.5109 (0.80)
	Φ_3	0.6454	0.5099 (0.65)	0.4962 (0.67)	0.6985	0.5358 (0.56)	0.5036 (0.59)
	Φ_4	0.5769	0.5108 (0.52)	0.4930 (0.53)	0.5503	0.5015 (0.55)	0.4828 (0.58)
	Φ_5	0.5304	0.5488 (0.30)	0.5273 (0.33)	0.5253	0.5859 (0.18)	0.5704 (0.18)
	Φ_6	0.5356	0.6493 (0.20)	0.6278 (0.21)	0.5047	0.6141 (0.22)	0.6006 (0.22)
100	Φ_1	0.4951	0.3356 (0.75)	0.3180 (0.81)	0.4820	0.3717 (0.63)	0.3423 (0.70)
	Φ_2	0.4397	0.3478 (0.65)	0.3339 (0.70)	0.4076	0.3204 (0.70)	0.3078 (0.74)
	Φ_3	0.4454	0.3817 (0.55)	0.3587 (0.61)	0.4138	0.3560 (0.54)	0.3375 (0.61)
	Φ_4	0.3437	0.3454 (0.36)	0.3327 (0.37)	0.3698	0.3548 (0.43)	0.3395 (0.48)
	Φ_5	0.3313	0.3742 (0.20)	0.3687 (0.22)	0.3569	0.3967 (0.18)	0.3847 (0.21)
	Φ_6	0.2961	0.4055 (0.14)	0.3977 (0.16)	0.2729	0.3902 (0.12)	0.3811 (0.15)
200	Φ_1	0.2761	0.1927 (0.61)	0.1791 (0.67)	0.2484	0.1878 (0.63)	0.1788 (0.64)
	Φ_2	0.2090	0.1645 (0.60)	0.1555 (0.65)	0.1870	0.1679 (0.43)	0.1582 (0.47)
	Φ_3	0.2008	0.2129 (0.36)	0.2043 (0.38)	0.1936	0.2212 (0.28)	0.2062 (0.33)
	Φ_4	0.1588	0.1814 (0.26)	0.1728 (0.30)	0.1824	0.2031 (0.30)	0.1981 (0.33)
	Φ_5	0.1585	0.2320 (0.13)	0.2293 (0.13)	0.1604	0.2376 (0.13)	0.2325 (0.13)
	Φ_6	0.1468	0.2461 (0.09)	0.2415 (0.04)	0.1432	0.2555 (0.04)	0.2526 (0.03)
1000	Φ_1	0.1243	0.0961 (0.55)	0.0905 (0.57)	0.1403	0.1070 (0.52)	0.1030 (0.54)
	Φ_2	0.1021	0.0969 (0.46)	0.0927 (0.47)	0.1020	0.0985 (0.38)	0.0942 (0.47)
	Φ_3	0.0962	0.1369 (0.22)	0.1324 (0.22)	0.1048	0.1345 (0.25)	0.1251 (0.25)
	Φ_4	0.0856	0.1308 (0.08)	0.1286 (0.10)	0.0890	0.1323 (0.14)	0.1303 (0.14)
	Φ_5	0.0852	0.1622 (0.05)	0.1599 (0.06)	0.0871	0.1701 (0.04)	0.1661 (0.02)
	Φ_6	0.0787	0.1723 (0.03)	0.1710 (0.02)	0.0739	0.1708 (0.02)	0.1679 (0.01)

Sample Size	Φ	Model 3 ML	Model 3 Lasso	Model 3 Lasso (OD)
50	Φ_1	0.8137	0.6029 (0.72)	0.5651 (0.84)
	Φ_2	0.7710	0.5883 (0.75)	0.5661 (0.82)
	Φ_3	0.6857	0.5798 (0.55)	0.5522 (0.63)
	Φ_4	0.6248	0.5520 (0.53)	0.5318 (0.55)
	Φ_5	0.6025	0.6246 (0.36)	0.6052 (0.39)
	Φ_6	0.5721	0.6654 (0.23)	0.6554 (0.23)
100	Φ_1	0.4689	0.3544 (0.72)	0.3363 (0.77)
	Φ_2	0.4624	0.3588 (0.73)	0.3523 (0.77)
	Φ_3	0.4129	0.3698 (0.50)	0.3521 (0.52)
	Φ_4	0.3660	0.3604 (0.47)	0.3413 (0.53)
	Φ_5	0.3667	0.4261 (0.24)	0.4063 (0.29)
	Φ_6	0.3210	0.4318 (0.15)	0.4312 (0.15)
200	Φ_1	0.2531	0.1915 (0.60)	0.1760 (0.67)
	Φ_2	0.2309	0.1771 (0.70)	0.1688 (0.71)
	Φ_3	0.1971	0.2379 (0.35)	0.2139 (0.40)
	Φ_4	0.1926	0.2145 (0.34)	0.2084 (0.34)
	Φ_5	0.1738	0.2682 (0.12)	0.2572 (0.12)
	Φ_6	0.1546	0.2642 (0.03)	0.2585 (0.04)
1000	Φ_1	0.1274	0.1033 (0.56)	0.0976 (0.60)
	Φ_2	0.1076	0.1003 (0.45)	0.0945 (0.50)
	Φ_3	0.1016	0.1499 (0.20)	0.1436 (0.20)
	Φ_4	0.0920	0.1367 (0.14)	0.1325 (0.14)
	Φ_5	0.0839	0.1703 (0.02)	0.1655 (0.01)
	Φ_6	0.0799	0.1745 (0.01)	0.1704 (0.02)

Table 6.15: Overall Discrepancy Table : Lasso SEM, No Model Error

OD : Overall Discrepancy (OLS)

Sample Size	Φ	Model 1 ML	Model 1 Lasso	Model 1 Lasso (OD)	Model 2 ML	Model 2 Lasso	Model 2 Lasso (OD)
50	Φ_1	0.8103	0.6379 (0.67)	0.6026 (0.75)	0.8621	0.6891 (0.66)	0.6502 (0.74)
	Φ_2	0.8108	0.6563 (0.73)	0.6368 (0.83)	0.7912	0.6885 (0.70)	0.6667 (0.72)
	Φ_3	0.7327	0.6226 (0.49)	0.5966 (0.54)	0.8256	0.6931 (0.49)	0.6561 (0.53)
	Φ_4	0.6943	0.6575 (0.51)	0.6248 (0.57)	0.6690	0.6531 (0.42)	0.6334 (0.52)
	Φ_5	0.6107	0.6781 (0.22)	0.6550 (0.22)	0.6509	0.7179 (0.28)	0.7074 (0.30)
	Φ_6	0.6091	0.7263 (0.19)	0.7081 (0.21)	0.6250	0.7562 (0.16)	0.7461 (0.16)
100	Φ_1	0.5676	0.4523 (0.63)	0.4188 (0.69)	0.5907	0.4518 (0.62)	0.4342 (0.68)
	Φ_2	0.5301	0.4533 (0.60)	0.4332 (0.68)	0.5047	0.4190 (0.57)	0.4059 (0.66)
	Φ_3	0.5470	0.4534 (0.44)	0.4395 (0.45)	0.4196	0.4175 (0.38)	0.3975 (0.44)
	Φ_4	0.4752	0.4510 (0.38)	0.4328 (0.41)	0.4778	0.4813 (0.30)	0.4678 (0.33)
	Φ_5	0.3901	0.4732 (0.12)	0.4616 (0.16)	0.3932	0.4847 (0.18)	0.4699 (0.18)
	Φ_6	0.4092	0.5074 (0.16)	0.4980 (0.17)	0.4131	0.5317 (0.14)	0.5268 (0.14)
200	Φ_1	0.2727	0.2198 (0.50)	0.2068 (0.51)	0.2804	0.2204 (0.53)	0.2074 (0.61)
	Φ_2	0.2676	0.2377 (0.54)	0.2276 (0.63)	0.2932	0.2574 (0.55)	0.2470 (0.58)
	Φ_3	0.2556	0.2700 (0.34)	0.2536 (0.39)	0.2653	0.2833 (0.30)	0.2707 (0.33)
	Φ_4	0.2640	0.2955 (0.27)	0.2907 (0.30)	0.2720	0.2996 (0.29)	0.2975 (0.36)
	Φ_5	0.2314	0.3126 (0.13)	0.2968 (0.15)	0.2430	0.3274 (0.13)	0.3172 (0.14)
	Φ_6	0.2390	0.3498 (0.06)	0.3463 (0.07)	0.2385	0.3710 (0.02)	0.3682 (0.04)
1000	Φ_1	0.1777	0.1499 (0.54)	0.1430 (0.54)	0.1660	0.1485 (0.48)	0.1395 (0.51)
	Φ_2	0.1817	0.1620 (0.52)	0.1571 (0.55)	0.1766	0.1650 (0.38)	0.1613 (0.43)
	Φ_3	0.1793	0.2002 (0.31)	0.1926 (0.35)	0.1619	0.2019 (0.22)	0.1965 (0.22)
	Φ_4	0.1809	0.2243 (0.23)	0.2218 (0.24)	0.1838	0.2420 (0.18)	0.2344 (0.19)
	Φ_5	0.1451	0.2295 (0.05)	0.2258 (0.04)	0.1585	0.2510 (0.05)	0.2457 (0.05)
	Φ_6	0.1561	0.2650 (0.01)	0.2636 (0.02)	0.1754	0.2858 (0.01)	0.2860 (0.01)

Sample Size	Φ	Model 3 ML	Model 3 Lasso	Model 3 Lasso (OD)
50	Φ_1	0.8910	0.6462 (0.71)	0.6130 (0.81)
	Φ_2	0.8139	0.6725 (0.72)	0.6599 (0.73)
	Φ_3	0.7631	0.6237 (0.49)	0.6021 (0.55)
	Φ_4	0.7096	0.6378 (0.49)	0.6188 (0.54)
	Φ_5	0.6447	0.6926 (0.33)	0.6721 (0.36)
	Φ_6	0.5294	0.6772 (0.13)	0.6597 (0.12)
100	Φ_1	0.5601	0.4051 (0.69)	0.3776 (0.78)
	Φ_2	0.4331	0.3544 (0.67)	0.3390 (0.74)
	Φ_3	0.3918	0.4018 (0.40)	0.3713 (0.44)
	Φ_4	0.3637	0.3843 (0.32)	0.3735 (0.37)
	Φ_5	0.3719	0.4912 (0.14)	0.4756 (0.16)
	Φ_6	0.3662	0.5096 (0.15)	0.5063 (0.16)
200	Φ_1	0.2616	0.2143 (0.51)	0.2004 (0.57)
	Φ_2	0.2539	0.2088 (0.59)	0.2008 (0.67)
	Φ_3	0.2163	0.2424 (0.28)	0.2348 (0.30)
	Φ_4	0.2239	0.2506 (0.33)	0.2427 (0.33)
	Φ_5	0.1938	0.3134 (0.11)	0.3023 (0.10)
	Φ_6	0.1746	0.3184 (0.02)	0.3169 (0.03)
1000	Φ_1	0.1459	0.1344 (0.46)	0.1272 (0.47)
	Φ_2	0.1189	0.1185 (0.37)	0.1148 (0.37)
	Φ_3	0.1183	0.1713 (0.19)	0.1683 (0.20)
	Φ_4	0.1159	0.1741 (0.13)	0.1680 (0.13)
	Φ_5	0.1014	0.2151 (0.03)	0.2144 (0.03)
	Φ_6	0.0918	0.2215 (0.01)	0.2202 (0.01)

Table 6.16: Overall Discrepancy Table : Lasso SEM, Model Error Involved

in the condition of Φ_5 and Φ_6. Φ_3 and Φ_4 yield varying results according to sample sizes and model error. Lasso performs better with small samples but as n increases, ML reverses the result. When the model error is employed, performance of both methods get worse which can be seen by increased OD mean estimates in table 6.16.

The degree of misspecification does not have considerable influence, but it affects moderately the results in Φ_3 and Φ_4. In general, overall tendency in OD comparison does not diverge that far from Lasso FA result, except that Φ_3 and Φ_4 conditions are less supportive for Lass.

A comparison study of sMSE also arrives at a similar conclusion to research 1. The sMSE estimates obtained by ML and Lasso are contained in table 6.17 and 6.18. For model 1 which has no misspecification, Lasso is no match for ML. However, Lasso yields great results as the degree of misspecification increases, especially for Φ_1 and Φ_2. This trend caused by misspecification is the notable distinction between sMSE and OD. Model error does not seem to greatly influence the outcomes in the matter of reducing sMSE.

In sum, the simulation results of L_1-Regularized SEM, regarding OD and sMSE, are similar to those of Lasso FA in that the Φ matrix turns out to be the dominant factor affecting the performances of estimation methods in reducing those indices which is supposed to represent the generalizability and reproducibility.

As Lasso FA results, DA, DE, bias squared and variance observations are provided in Subsection 6.3.1 and Appendix A. Regarding DA and DE, we arrived the same conclusion with that in research 1; Lasso SEM also enjoys remarkably small DE but suffers from fairly large DA, which offsets the advantage and hinders Lasso SEM from producing better results.

sMSE : Mean Squared Error of Standardized Parameter Estimates

Sample Size	Φ	Model 1 ML	Model 1 Lasso	Model 1 Lasso (OD)	Model 2 ML	Model 2 Lasso	Model 2 Lasso (OD)
50	Φ_1	0.1005	0.1061 (0.51)	0.1018 (0.60)	0.1251	0.1262 (0.59)	0.1230 (0.61)
	Φ_2	0.1116	0.1191 (0.47)	0.1176 (0.46)	0.1378	0.1404 (0.54)	0.1410 (0.51)
	Φ_3	0.1637	0.1782 (0.42)	0.1786 (0.44)	0.2157	0.2165 (0.40)	0.2100 (0.43)
	Φ_4	0.1819	0.2162 (0.27)	0.2128 (0.26)	0.2210	0.2435 (0.33)	0.2424 (0.31)
	Φ_5	0.2504	0.3355 (0.17)	0.3343 (0.17)	0.4498	0.4290 (0.17)	0.4223 (0.18)
	Φ_6	0.3051	0.4822 (0.07)	0.4779 (0.06)	0.3946	0.4889 (0.21)	0.4932 (0.20)
100	Φ_1	0.0543	0.0571 (0.46)	0.0569 (0.48)	0.0746	0.0823 (0.45)	0.0784 (0.51)
	Φ_2	0.0671	0.0737 (0.41)	0.0722 (0.39)	0.0848	0.0845 (0.50)	0.0829 (0.55)
	Φ_3	0.0938	0.1161 (0.23)	0.1127 (0.24)	0.1251	0.1370 (0.33)	0.1329 (0.38)
	Φ_4	0.1027	0.1370 (0.08)	0.1356 (0.09)	0.1377	0.1588 (0.35)	0.1550 (0.32)
	Φ_5	0.1513	0.2241 (0.13)	0.2231 (0.12)	0.1992	0.2762 (0.21)	0.2692 (0.19)
	Φ_6	0.1699	0.2959 (0.02)	0.2966 (0.02)	0.1924	0.3143 (0.06)	0.3133 (0.08)
200	Φ_1	0.0278	0.0304 (0.44)	0.0295 (0.41)	0.0342	0.0354 (0.42)	0.0343 (0.48)
	Φ_2	0.0318	0.0340 (0.42)	0.0335 (0.42)	0.0380	0.0403 (0.39)	0.0397 (0.34)
	Φ_3	0.0469	0.0648 (0.11)	0.0637 (0.12)	0.0579	0.0772 (0.17)	0.0751 (0.18)
	Φ_4	0.0524	0.0709 (0.13)	0.0694 (0.12)	0.0615	0.0827 (0.14)	0.0820 (0.14)
	Φ_5	0.0745	0.1387 (0.00)	0.1385 (0.01)	0.0965	0.1641 (0.05)	0.1614 (0.07)
	Φ_6	0.0858	0.1800 (0.01)	0.1778 (0.01)	0.0987	0.1988 (0.04)	0.1994 (0.04)
1000	Φ_1	0.0133	0.0150 (0.37)	0.0147 (0.39)	0.0174	0.0199 (0.44)	0.0197 (0.45)
	Φ_2	0.0159	0.0189 (0.28)	0.0186 (0.30)	0.0197	0.0228 (0.34)	0.0224 (0.32)
	Φ_3	0.0229	0.0361 (0.03)	0.0360 (0.03)	0.0277	0.0418 (0.17)	0.0399 (0.19)
	Φ_4	0.0265	0.0443 (0.05)	0.0443 (0.04)	0.0310	0.0491 (0.08)	0.0490 (0.08)
	Φ_5	0.0369	0.0856 (0.00)	0.0851 (0.00)	0.0509	0.1023 (0.07)	0.1014 (0.08)
	Φ_6	0.0417	0.1250 (0.00)	0.1246 (0.00)	0.0468	0.1264 (0.01)	0.1251 (0.01)

Sample Size	Φ	Model 3 ML	Model 3 Lasso	Model 3 Lasso (OD)
50	Φ_1	0.2104	0.1880 (0.65)	0.1843 (0.74)
	Φ_2	0.1934	0.1831 (0.66)	0.1802 (0.66)
	Φ_3	0.3858	0.2966 (0.59)	0.2939 (0.65)
	Φ_4	0.2695	0.2699 (0.45)	0.2670 (0.46)
	Φ_5	1.1297	0.5864 (0.40)	0.5907 (0.41)
	Φ_6	0.4809	0.5805 (0.17)	0.5540 (0.19)
100	Φ_1	0.1204	0.1044 (0.74)	0.1032 (0.73)
	Φ_2	0.1149	0.1055 (0.58)	0.1053 (0.59)
	Φ_3	0.2041	0.1799 (0.51)	0.1786 (0.52)
	Φ_4	0.1706	0.1740 (0.47)	0.1679 (0.50)
	Φ_5	0.4668	0.4498 (0.33)	0.4373 (0.33)
	Φ_6	0.2605	0.3716 (0.15)	0.3739 (0.14)
200	Φ_1	0.0548	0.0470 (0.74)	0.0461 (0.79)
	Φ_2	0.0543	0.0512 (0.56)	0.0501 (0.60)
	Φ_3	0.1039	0.0980 (0.47)	0.0941 (0.48)
	Φ_4	0.0845	0.0919 (0.30)	0.0909 (0.28)
	Φ_5	0.1785	0.2118 (0.24)	0.2087 (0.24)
	Φ_6	0.1263	0.2181 (0.05)	0.2159 (0.05)
1000	Φ_1	0.0288	0.0246 (0.71)	0.0245 (0.70)
	Φ_2	0.0267	0.0260 (0.50)	0.0252 (0.51)
	Φ_3	0.0446	0.0542 (0.25)	0.0536 (0.28)
	Φ_4	0.0369	0.0523 (0.12)	0.0517 (0.13)
	Φ_5	0.0842	0.1225 (0.20)	0.1206 (0.21)
	Φ_6	0.0606	0.1355 (0.04)	0.1334 (0.04)

Table 6.17: Standardized Mean Squared Error Table - Lasso SEM, No Model Error

sMSE : Mean Squared Error of Standardized Parameter Estimates

Sample Size	Φ	Model 1			Model 2		
		ML	Lasso	Lasso (OD)	ML	Lasso	Lasso (OD)
50	Φ_1	0.0952	0.0965 (0.61)	0.0950 (0.61)	0.1299	0.1339 (0.54)	0.1301 (0.56)
	Φ_2	0.1170	0.1181 (0.60)	0.1162 (0.63)	0.1481	0.1565 (0.53)	0.1541 (0.57)
	Φ_3	0.1571	0.1678 (0.29)	0.1659 (0.33)	0.2193	0.2386 (0.39)	0.2333 (0.39)
	Φ_4	0.1821	0.1969 (0.32)	0.1916 (0.35)	0.2320	0.2509 (0.24)	0.2482 (0.27)
	Φ_5	0.2497	0.3440 (0.12)	0.3444 (0.13)	1.7742	0.4728 (0.25)	0.4649 (0.27)
	Φ_6	0.2945	0.4115 (0.14)	0.4092 (0.13)	0.3685	0.4997 (0.14)	0.4954 (0.14)
100	Φ_1	0.0565	0.0554 (0.57)	0.0536 (0.61)	0.0735	0.0746 (0.56)	0.0739 (0.56)
	Φ_2	0.0675	0.0706 (0.51)	0.0686 (0.52)	0.0892	0.0941 (0.40)	0.0931 (0.46)
	Φ_3	0.1028	0.1090 (0.34)	0.1083 (0.34)	0.1373	0.1499 (0.33)	0.1480 (0.35)
	Φ_4	0.1101	0.1226 (0.26)	0.1199 (0.27)	0.1464	0.1707 (0.22)	0.1676 (0.22)
	Φ_5	0.1458	0.2229 (0.03)	0.2228 (0.02)	0.2317	0.3167 (0.19)	0.3125 (0.20)
	Φ_6	0.1661	0.2662 (0.06)	0.2678 (0.07)	0.2358	0.3504 (0.06)	0.3504 (0.07)
200	Φ_1	0.0268	0.0271 (0.42)	0.0268 (0.48)	0.0394	0.0400 (0.43)	0.0393 (0.49)
	Φ_2	0.0324	0.0326 (0.47)	0.0321 (0.46)	0.0498	0.0517 (0.45)	0.0509 (0.49)
	Φ_3	0.0449	0.0537 (0.23)	0.0528 (0.25)	0.0713	0.0870 (0.12)	0.0859 (0.18)
	Φ_4	0.0522	0.0645 (0.17)	0.0642 (0.18)	0.0803	0.0946 (0.09)	0.0949 (0.08)
	Φ_5	0.0722	0.1281 (0.08)	0.1235 (0.06)	0.1329	0.1907 (0.16)	0.1870 (0.15)
	Φ_6	0.0832	0.1616 (0.02)	0.1607 (0.02)	0.1237	0.2188 (0.04)	0.2183 (0.03)
1000	Φ_1	0.0133	0.0138 (0.46)	0.0133 (0.47)	0.0222	0.0238 (0.34)	0.0235 (0.38)
	Φ_2	0.0161	0.0165 (0.37)	0.0162 (0.39)	0.0319	0.0334 (0.28)	0.0332 (0.31)
	Φ_3	0.0228	0.0308 (0.14)	0.0305 (0.13)	0.0440	0.0576 (0.04)	0.0574 (0.04)
	Φ_4	0.0256	0.0364 (0.07)	0.0360 (0.07)	0.0567	0.0727 (0.03)	0.0714 (0.03)
	Φ_5	0.0350	0.0747 (0.03)	0.0741 (0.02)	0.0774	0.1354 (0.02)	0.1346 (0.02)
	Φ_6	0.0410	0.0992 (0.02)	0.0990 (0.02)	0.0794	0.1471 (0.01)	0.1472 (0.01)

Sample Size	Φ	Model 3		
		ML	Lasso	Lasso (OD)
50	Φ_1	0.2552	0.2571 (0.55)	0.2532 (0.65)
	Φ_2	0.2329	0.2313 (0.66)	0.2290 (0.67)
	Φ_3	0.4938	0.5061 (0.48)	0.5037 (0.47)
	Φ_4	0.4309	0.4295 (0.44)	0.4273 (0.43)
	Φ_5	0.9786	0.8748 (0.26)	0.8723 (0.27)
	Φ_6	0.6068	0.7014 (0.14)	0.6931 (0.16)
100	Φ_1	0.1925	0.1851 (0.69)	0.1831 (0.67)
	Φ_2	0.1491	0.1448 (0.61)	0.1450 (0.60)
	Φ_3	0.3658	0.3899 (0.18)	0.3864 (0.20)
	Φ_4	0.2566	0.2632 (0.34)	0.2624 (0.33)
	Φ_5	0.9688	0.7152 (0.19)	0.7207 (0.20)
	Φ_6	0.4460	0.5539 (0.09)	0.5528 (0.11)
200	Φ_1	0.1199	0.1194 (0.52)	0.1187 (0.59)
	Φ_2	0.0935	0.0912 (0.68)	0.0903 (0.68)
	Φ_3	0.2675	0.2905 (0.10)	0.2901 (0.12)
	Φ_4	0.1925	0.1973 (0.34)	0.1964 (0.35)
	Φ_5	0.4915	0.5747 (0.11)	0.5667 (0.11)
	Φ_6	0.3078	0.3738 (0.16)	0.3745 (0.16)
1000	Φ_1	0.0912	0.0922 (0.39)	0.0919 (0.46)
	Φ_2	0.0674	0.0657 (0.66)	0.0652 (0.74)
	Φ_3	0.2189	0.2449 (0.02)	0.2447 (0.02)
	Φ_4	0.1385	0.1425 (0.36)	0.1420 (0.37)
	Φ_5	0.3860	0.4610 (0.04)	0.4601 (0.05)
	Φ_6	0.2287	0.2790 (0.04)	0.2781 (0.05)

Table 6.18: Standardized Mean Squared Error Table - Lasso SEM, Model Error Involved

The following tables and figures represent the estimates of parameters in SEM. Note that we select and post the result obtained from the Φ_4 condition for the same reason mentioned in the previous section. And the plots are computed using the second sample size condition, namely $n = 250$ in SEM.

In this investigation, we mainly focus on comparing the misspecified parameters with correctly specified ones. Starred coefficients in tables and figures indicate the misspecified paths. Note that in research 2, we only study on the misspecification for the structural model, not for the measurement model. However, the overall results are similar to factor loading estimates and their plots in research 1. In the figure, each of rows contains the distributions of correctly specified and misspecified parameters, respectively. If we assume no model error in population, Lasso yields sharp distributions, whose modes are nearly equal to zero, for the nuisance parameters as presented in the second row. It seems that these results are caused by Lasso's complete shrinkage property. Correctly specified parameters also have more narrow distributions than ML, concentrated closely on the true parameter values. Hence, it can be said that the proposed Lasso SEM algorithm works fairly well when the population can be assumed to be largely generated from the underlying systematic process.

As in the Lasso FA, however, the model error brings about a considerable change in the outcome. At first, estimates decrease for correctly specified regression coefficients such as $\beta_1, \beta_2, \gamma_1$, and γ_2. On the contrary, misspecified parameters shows larger estimates when the model error is involved. Nonetheless, note that these changes on path coefficients are common in both ML and Lasso. One noticeable difference between those estimation methods is that Lasso produces larger standard deviations for all the parameters whether they are misspecified or not. However, these variabilities are still less than ML's

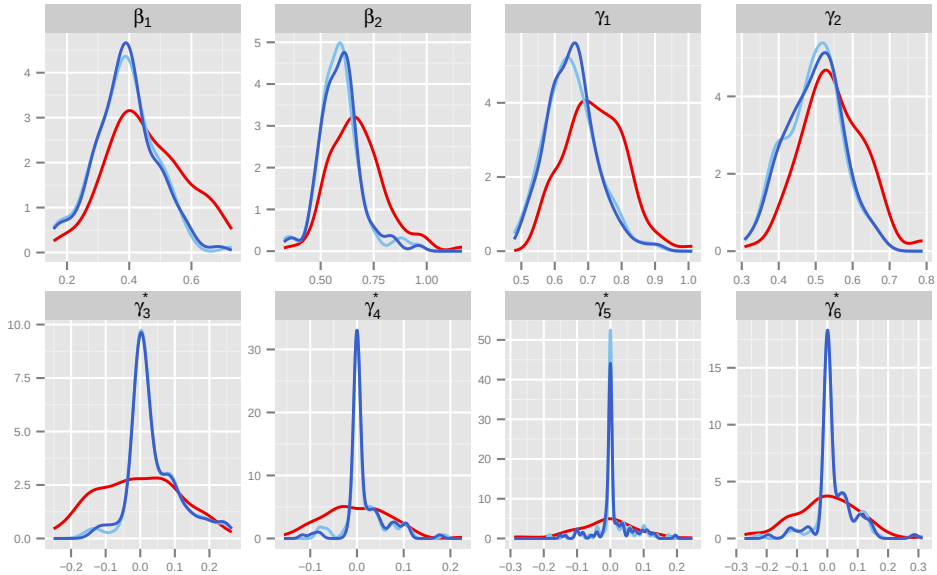

Figure 6.11: Density Plots for Selected Parameters - SEM, Model 3, No Model Error, $N = 250$, Φ_4

		Sample Size = 150		*Sample Size = 250*		*Sample Size = 500*		*Sample Size = 1,000*	
	True	**ML**	**LASSO**	**ML**	**LASSO**	**ML**	**LASSO**	**ML**	**LASSO**
β_1	*0.45*	0.4558	0.3880	0.4554	0.3876	0.4534	0.3997	0.4399	0.4004
		(0.1282)	(0.1167)	(0.1263)	(0.0997)	(0.0764)	(0.0588)	(0.0518)	(0.0419)
β_2	*0.65*	0.6390	0.5648	0.6735	0.5910	0.6478	0.5825	0.6512	0.5900
		(0.1471)	(0.1004)	(0.1311)	(0.0967)	(0.0927)	(0.0650)	(0.0496)	(0.0386)
γ_1	*0.70*	0.7048	0.6307	0.7197	0.6504	0.7001	0.6420	0.7035	0.6563
		(0.1399)	(0.1164)	(0.0890)	(0.0749)	(0.0656)	(0.0570)	(0.0453)	(0.0413)
γ_2	*0.55*	0.5626	0.5132	0.5468	0.4956	0.5551	0.5112	0.5456	0.5102
		(0.1050)	(0.1009)	(0.0850)	(0.0749)	(0.0643)	(0.0570)	(0.0380)	(0.0346)
γ_3^*	-	**-0.0199**	**0.0249**	**-0.0057**	**0.0395**	**-0.0052**	**0.0289**	**0.0076**	**0.0241**
		(0.1345)	**(0.0909)**	**(0.1178)**	**(0.0754)**	**(0.0840)**	**(0.0518)**	**(0.0487)**	**(0.0314)**
γ_4^*	-	**0.0088**	**0.0230**	**-0.0025**	**0.0145**	**0.0066**	**0.0156**	**0.0028**	**0.0057**
		(0.1000)	**(0.0612)**	**(0.0700)**	**(0.0413)**	**(0.0632)**	**(0.0340)**	**(0.0362)**	**(0.0147)**
γ_5^*	-	**0.0174**	**0.0252**	**-0.0130**	**0.0087**	**0.0018**	**0.0080**	**-0.0020**	**0.0056**
		(0.1044)	**(0.0575)**	**(0.0948)**	**(0.0519)**	**(0.0528)**	**(0.0231)**	**(0.0409)**	**(0.0140)**
γ_6^*	-	**0.0143**	**0.0346**	**-0.0026**	**0.0192**	**-0.0004**	**0.0143**	**-0.0006**	**0.0096**
		(0.1317)	**(0.0848)**	**(0.1070)**	**(0.0635)**	**(0.0740)**	**(0.0355)**	**(0.0463)**	**(0.0187)**
$\psi^2_{\zeta 1}$	*0.30*	0.2976	0.3288	0.2945	0.3251	0.2992	0.3272	0.3010	0.3282
		(0.0760)	(0.0887)	(0.0587)	(0.0676)	(0.0411)	(0.0466)	(0.0305)	(0.0363)
$\psi^2_{\zeta 2}$	*0.25*	0.2421	0.2712	0.2367	0.2687	0.2461	0.2747	0.2513	0.2770
		(0.0717)	(0.0785)	(0.0444)	(0.0504)	(0.0404)	(0.0446)	(0.0279)	(0.0309)
$\psi^2_{\zeta 3}$	*0.28*	0.2726	0.2883	0.2638	0.2800	0.2743	0.2866	0.2816	0.2916
		(0.0773)	(0.0873)	(0.0538)	(0.0591)	(0.0378)	(0.0419)	(0.0277)	(0.0297)

Table 6.19: Estimates of Parameters - SEM, Model 3, No Model Error, Φ_4

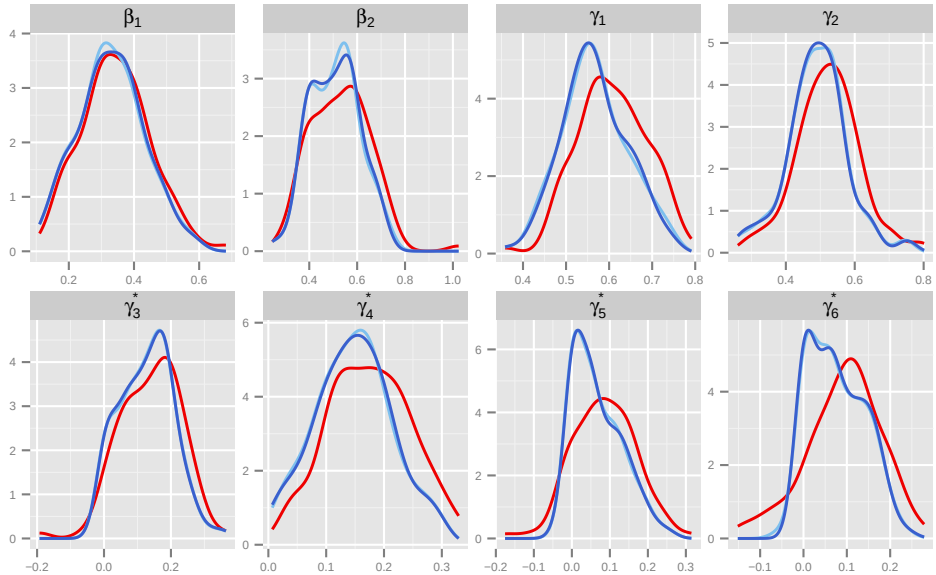

Figure 6.12: Density Plots for Selected Parameters - SEM, Model 3, Model Error Involved, $N = 250$, Φ_4

		Sample Size = 150		*Sample Size = 250*		*Sample Size = 500*		*Sample Size = 1,000*	
	True	**ML**	**LASSO**	**ML**	**LASSO**	**ML**	**LASSO**	**ML**	**LASSO**
β_1	*0.45*	0.3203 (0.1480)	0.2960 (0.1383)	0.3448 (0.1066)	0.3294 (0.1017)	0.3210 (0.0781)	0.3082 (0.0725)	0.3312 (0.0489)	0.3213 (0.0481)
β_2	*0.65*	0.5447 (0.2023)	0.4968 (0.1431)	0.5312 (0.1242)	0.5062 (0.1017)	0.5294 (0.0854)	0.5138 (0.0766)	0.5461 (0.0569)	0.5348 (0.0510)
γ_1	*0.70*	0.5893 (0.1187)	0.5459 (0.1164)	0.6072 (0.0806)	0.5662 (0.0757)	0.6129 (0.0487)	0.5762 (0.0461)	0.6183 (0.0423)	0.5828 (0.0403)
γ_2	*0.55*	0.5197 (0.1121)	0.4834 (0.1034)	0.5228 (0.0929)	0.4905 (0.0883)	0.5302 (0.0607)	0.5007 (0.0585)	0.5243 (0.0400)	0.4973 (0.0391)
γ_3^*	-	**0.1535 (0.1470)**	**0.1441 (0.1317)**	**0.1393 (0.0909)**	**0.1272 (0.0795)**	**0.1361 (0.0753)**	**0.1215 (0.0723)**	**0.1409 (0.0548)**	**0.1268 (0.0504)**
γ_4^*	-	**0.2000 (0.1001)**	**0.1656 (0.0927)**	**0.1755 (0.0706)**	**0.1470 (0.0667)**	**0.1832 (0.0550)**	**0.1537 (0.0509)**	**0.1763 (0.0390)**	**0.1492 (0.0372)**
γ_5^*	-	**0.0776 (0.1261)**	**0.0704 (0.0956)**	**0.0851 (0.0828)**	**0.0665 (0.0647)**	**0.0884 (0.0607)**	**0.0667 (0.0528)**	**0.0765 (0.0369)**	**0.0525 (0.0309)**
γ_6^*	-	**0.0907 (0.1615)**	**0.0871 (0.1055)**	**0.0916 (0.0864)**	**0.0763 (0.0632)**	**0.0922 (0.0775)**	**0.0695 (0.0656)**	**0.0794 (0.0521)**	**0.0565 (0.0439)**
$\psi^2_{\zeta 1}$	*0.30*	0.2970 (0.0769)	0.3174 (0.0863)	0.3024 (0.0579)	0.3233 (0.0635)	0.3060 (0.0393)	0.3285 (0.0449)	0.3096 (0.0288)	0.3308 (0.0326)
$\psi^2_{\zeta 2}$	*0.25*	0.2653 (0.0770)	0.2891 (0.0807)	0.2674 (0.0522)	0.2868 (0.0557)	0.2730 (0.0350)	0.2917 (0.0385)	0.2668 (0.0285)	0.2835 (0.0310)
$\psi^2_{\zeta 3}$	*0.28*	0.2700 (0.0622)	0.2829 (0.0657)	0.2778 (0.0553)	0.2849 (0.0588)	0.2873 (0.0408)	0.2946 (0.0455)	0.2860 (0.0267)	0.2915 (0.0289)

Table 6.20: Estimates of Parameters - SEM, Model 3, Model Error Involved, Φ_4

standard deviation even though the model error increases them. Perturbations ψ_{ζ}^{2}'s tend to be estimated larger in Lasso than the existing method. This is reasonable since the regression coefficients are generally shrunk toward zero so that have smaller values in Lasso results. This decreases the proportion of variabilities of η explained by their regressors in the structural model, which leaves larger proportions supposed to belong to ζ.

Density plots reflects these changes. Almost all parameters, including correctly specified ones, still have more dense distributions than ML outcomes. However, the differences are largely wiped out. For misspecified parameters, it is hard to say that densities are centered at zero any more.

At last, we study on Lasso's capability to shrink and remove some estimates of nuisance parameters. Table 6.21, and table 6.22 $\sim$ 6.23 provide us the information required to conduct the study. At first, in table 6.21, model error acts in the similar way as in the Lasso FA. Without model error, Lasso is able to exert its ability adequately. The proportion of shrinkage occurrence is considerably high over all types of Φ. The effect appears more satisfactorily for γ_5 and γ_6 than γ_3 and γ_4. With model error, however, the result becomes worse to the same extent as in the factor analysis model. Moreover, the contribution of increasing sample sizes supports the interpretation mentioned in investigation of Lasso FA; occurrence of shrinkage becomes more frequent as n gets larger when the model error is not employed, while the error seems to be mistaken for meaningful variability with large sample size in model error-involved case.

Another noteworthy point is that, unlike the result of research 1, the shrinkage trends do not show any remarkable differences between Φ_3, Φ_5, and the other conditions. This seems due to the fact than SEM contains both of endogenous and exogenous latent variables. The covariance matrix among latent variables contains Φ as its right-below submatrix in SEM. But in FA, Φ per se

Φ		Without Model Error								With Model Error							
		LASSO				LASSO.OD				LASSO				LASSO.OD			
		150	*250*	*500*	*1000*	*150*	*250*	*500*	*1000*	*150*	*250*	*500*	*1000*	*150*	*250*	*500*	*1000*
Φ_1	γ_3	0.18	0.23	0.30	0.34	0.20	0.27	0.31	0.31	0.11	0.16	0.09	0.04	0.13	0.14	0.08	0.04
	γ_4	0.36	0.25	0.37	0.42	0.35	0.28	0.35	0.39	0.05	0.00	0.00	0.00	0.04	0.00	0.00	0.00
	γ_5	0.37	0.36	0.43	0.54	0.38	0.44	0.39	0.52	0.27	0.30	0.16	0.16	0.24	0.31	0.15	0.18
	γ_6	0.34	0.33	0.43	0.48	0.33	0.39	0.43	0.49	0.20	0.26	0.22	0.26	0.20	0.26	0.22	0.24
Φ_2	γ_3	0.29	0.37	0.26	0.34	0.26	0.41	0.27	0.35	0.11	0.04	0.06	0.02	0.13	0.04	0.07	0.02
	γ_4	0.39	0.43	0.34	0.49	0.32	0.44	0.34	0.49	0.08	0.01	0.01	0.00	0.07	0.01	0.02	0.00
	γ_5	0.31	0.39	0.46	0.52	0.34	0.42	0.46	0.54	0.17	0.23	0.21	0.11	0.18	0.22	0.20	0.11
	γ_6	0.34	0.37	0.42	0.46	0.35	0.39	0.42	0.48	0.21	0.29	0.18	0.20	0.23	0.26	0.23	0.21
Φ_3	γ_3	0.30	0.26	0.38	0.19	0.30	0.29	0.38	0.21	0.10	0.08	0.07	0.02	0.09	0.08	0.08	0.02
	γ_4	0.41	0.38	0.42	0.48	0.39	0.37	0.43	0.47	0.02	0.01	0.00	0.00	0.03	0.01	0.00	0.00
	γ_5	0.44	0.46	0.67	0.68	0.45	0.52	0.71	0.69	0.23	0.25	0.19	0.13	0.25	0.24	0.17	0.14
	γ_6	0.42	0.53	0.67	0.71	0.43	0.51	0.66	0.70	0.33	0.23	0.18	0.14	0.28	0.23	0.19	0.14
Φ_4	γ_3	0.36	0.39	0.44	0.37	0.34	0.37	0.46	0.37	0.15	0.07	0.05	0.01	0.12	0.07	0.05	0.01
	γ_4	0.38	0.46	0.50	0.69	0.34	0.49	0.48	0.69	0.02	0.00	0.01	0.00	0.02	0.00	0.01	0.00
	γ_5	0.38	0.53	0.59	0.70	0.34	0.48	0.58	0.71	0.28	0.23	0.14	0.08	0.31	0.23	0.14	0.08
	γ_6	0.36	0.48	0.51	0.65	0.34	0.45	0.50	0.63	0.19	0.16	0.19	0.13	0.19	0.18	0.19	0.12
Φ_5	γ_3	0.34	0.34	0.35	0.26	0.36	0.33	0.29	0.26	0.17	0.11	0.08	0.02	0.17	0.12	0.08	0.02
	γ_4	0.32	0.35	0.38	0.34	0.30	0.34	0.35	0.31	0.12	0.04	0.01	0.00	0.09	0.02	0.00	0.00
	γ_5	0.59	0.61	0.79	0.78	0.61	0.59	0.81	0.79	0.26	0.28	0.22	0.09	0.27	0.28	0.20	0.11
	γ_6	0.56	0.69	0.67	0.80	0.55	0.65	0.72	0.82	0.36	0.32	0.19	0.09	0.35	0.33	0.14	0.10
Φ_6	γ_3	0.38	0.35	0.45	0.45	0.39	0.35	0.45	0.46	0.25	0.14	0.07	0.00	0.24	0.14	0.08	0.00
	γ_4	0.44	0.50	0.51	0.59	0.46	0.48	0.52	0.60	0.00	0.01	0.00	0.00	0.01	0.01	0.00	0.00
	γ_5	0.52	0.60	0.66	0.71	0.56	0.61	0.69	0.73	0.24	0.20	0.19	0.09	0.27	0.20	0.17	0.08
	γ_6	0.49	0.54	0.60	0.70	0.47	0.54	0.61	0.69	0.26	0.16	0.15	0.10	0.27	0.16	0.16	0.10

Table 6.21: Proportion of Complete Shrinkage to Zero - Model 3

is the latent variable covariance matrix. Therefore, relative magnitudes of variance and covariance in Φ dominates the covariance structure in FA, whereas SEM has other components that influence relationships among the parame-

ters. Hence the ratio of off-diagonal to on-diagonal elements of Φ, which is one of the determinative on occurrence of complete shrinkage in FA, has only the limited effect in SEM.

What seems important in Lasso SEM is the overall magnitudes of Φ. Φ_1 and Φ_2, which contains larger values, produce slightly smaller proportions than the other ϕ matrices. Note that these conditions make the measurement variables highly correlated in SEM, even though some of them are belonging to different factors. These high correlation may disrupt Lasso detecting and removing the misspecified parameters, since they show certain variabilities which can be misunderstood as reflecting meaningful relationships.

Table 6.22 and 6.23 display the outcome acquired by applying the strategy using a string of bounds in judging whether the estimates can be considered as zero or not. The result is not different from Lasso FA outcome; it is corroborated that the proposed Lasso SEM algorithm is able to yield complete shrinkage effect, but it suffers from the model error.

Table 6.21 $\sim$ 6.23 demonstrate again that model error throws a gloom over our expectation on Lasso SEM's shrinkage ability. As mentioned in Lasso FA result, However, some modification of Lasso SEM, which will be proposed in the Discussion chapter, seems to be able to shed light on this issue.

Shrinakge Effect in Lasso SEM : Without Model Error

	Model 2				*Model 3*			
	Total Number of Trials		2,400		Total Number of Trials		9,600	
	Number of Non-convergence		1		Number of Non-convergence		12	
	Number of Shrinkage Cases		2,399		Number of Shrinkage Cases		9,588	
Bound	**ML**		**Lasso**		**ML**		**Lasso**	
10^{-1}	1,547	(64.5%)	1,845	(76.9%)	6,702	(69.9%)	8,285	(86.4%)
10^{-2}	220	(9.17%)	942	(39.3%)	915	(9.54%)	5,002	(52.2%)
10^{-3}	15	(0.63%)	799	(33.3%)	87	(0.91%)	4,456	(46.5%)
10^{-4}	4	(0.17%)	783	(32.6%)	11	(0.11%)	4,400	(45.9%)
10^{-5}	-	-	782	(32.6%)	4	(0.04%)	4,391	(45.8%)
⋮		⋮				⋮		
10^{-15}	-	-	782	(32.6%)	-	-	4,388	(45.8%)

Table 6.22: Trends of Shrinkage Effect in Lasso FA - Without Model Error

Shrinakge Effect in Lasso SEM : With Model Error δ

	Model 2				*Model 3*			
	Total Number of Trials		2,400		Total Number of Trials		9,600	
	Number of Non-convergence		-		Number of Non-convergence		-	
	Number of Shrinkage Cases		2,400		Number of Shrinkage Cases		9,600	
Bound	**ML**		**Lasso**		**ML**		**Lasso**	
10^{-1}	1,027	(42.8%)	1,095	(45.6%)	3,866	(40.3%)	4,870	(50.7%)
10^{-2}	105	(4.38%)	290	(12.1%)	396	(4.13%)	1,525	(15.9%)
10^{-3}	17	(0.71%)	228	(9.50%)	41	(0.43%)	1,240	(12.9%)
10^{-4}	2	(0.08%)	220	(9.17%)	3	(0.03%)	1,212	(12.6%)
10^{-5}	-	-	218	(9.08%)	-	-	1,208	(12.6%)
⋮		⋮				⋮		
10^{-15}	-	-	218	(9.08%)	-	-	1,207	(12.6%)

Table 6.23: Trends of Shrinkage Effect in Lasso FA - With Model Error δ

6.3 Research 3: Additional Analyses

In this section, we add more information that provides answers to questions in Chapter 5. Also, some of the outcomes may give insight into the meaningful understanding of research results obtained in previous sections.

6.3.1 DA and Bias Analysis

In the main report on the results, our focus was on OD and sMSE, which can be viewed as generalizability indices. As described in Chapter 1 and 2, OD is approximately equal to the sum of DA and DE, and MSE can be decomposed into bias squared and variance. Therefore, investigating those components may help us elucidate the results related to generalizability and reproducibility of the analysis outcomes.

DE, which is parallel to variance or sampling error, shows far better results than OD does; Lasso SEM outperforms ML in reducing DE for almost all conditions. Thus, DA seems to be the reason that yields less supportive OD outcomes by getting too bad in Lasso estimation. Even though they are not good enough as DE results are, outcomes of variance tables also support Lasso SEM in some of their conditions with consistency. Research conditions where MSE and variance have different tendencies can be an issue, encouraging us to scrutinize bias of the outcome.

In this regard, we shall study on DA and bias squared tables in order to compare their values between Lasso SEM and ML. By doing so, we expect to attain some intuition about how we can understand the observed phenomena and how we can improve Lasso SEM's performance.

Tables of DE and variance are also presented in Appendix A for additional information.

1) Discrepancy due to Approximation

Table 6.24 and 6.25 show DA estimates obtained from the Lasso FA/SEM simulations when the model error is not employed. The following two tables Tables, 6.26 and 6.27, deal with the outcome from the case we include the model error in data-generating. In all the tables, some cells are filled with '10e-α' indicating that the observed values corresponding those cells are less than 10e-α.

The results can be abbreviated as follows.

- Over almost all conditions and models, ML yields the smallest DA and bias.
- Lasso suffers from large DA values, especially in Φ_5 and Φ_6 conditions. This can be one of the causes that deteriorates Lasso's performance on reducing OD in those conditions.
- In table 6.26 and 6.27, where the analyses contain model error, we can observe that model 1's in both FA and SEM recover the model error δ as its DA value(corresponding values are bolded). Since we insert the error based on ML-type discrepancy function, this can be observed at ML columns.

DA : Discrepancy due to Approximation (FA)

Estimation Method	Φ	Model 1			Model 2		
		ML	GLS	OLS	ML	GLS	OLS
ML	Φ_1	10e-8	10e-8	10e-8	10e-8	10e-8	10e-8
	Φ_2	10e-8	10e-8	10e-8	10e-8	10e-8	10e-8
	Φ_3	10e-8	10e-8	10e-8	10e-8	10e-8	10e-8
	Φ_4	10e-8	10e-8	10e-8	10e-8	10e-8	10e-8
	Φ_5	10e-8	10e-8	10e-8	10e-8	10e-8	10e-8
	Φ_6	10e-8	10e-8	10e-8	10e-8	10e-8	10e-8
Lasso	Φ_1	0.006224	0.006027	0.004961	0.006819	0.006588	0.005209
	Φ_2	0.005769	0.005602	0.004126	0.005771	0.005603	0.004125
	Φ_3	0.028090	0.025229	0.034530	0.030040	0.026954	0.034515
	Φ_4	0.009492	0.009441	0.016127	0.009491	0.009441	0.016127
	Φ_5	0.064267	0.053608	0.099351	0.067004	0.056013	0.100197
	Φ_6	0.020857	0.020597	0.040001	0.020857	0.020597	0.040001
Estimation Method	Φ	Model 3			Model 4		
		ML	GLS	OLS	ML	GLS	OLS
ML	Φ_1	10e-8	10e-8	10e-8	10e-8	10e-8	10e-8
	Φ_2	10e-8	10e-8	10e-8	10e-8	10e-8	10e-8
	Φ_3	10e-8	10e-8	10e-8	10e-8	10e-8	10e-8
	Φ_4	10e-8	10e-8	10e-8	10e-8	10e-8	10e-8
	Φ_5	10e-8	10e-8	10e-8	10e-8	10e-8	10e-8
	Φ_6	10e-8	10e-8	10e-8	10e-8	10e-8	10e-8
Lasso	Φ_1	0.007230	0.006978	0.004836	0.006274	0.006074	0.004943
	Φ_2	0.005771	0.005603	0.004125	0.005820	0.005649	0.004109
	Φ_3	0.031734	0.028368	0.029581	0.028067	0.025205	0.034504
	Φ_4	0.009491	0.009441	0.016127	0.009667	0.009643	0.016180
	Φ_5	0.070353	0.058793	0.092351	0.064252	0.053593	0.099327
	Φ_6	0.020857	0.020597	0.040001	0.020844	0.020583	0.039986

Table 6.24: DA Table - FA, No Model Error

DA : Discrepancy due to Approximation (SEM)

Estimation Method	Φ	Model 1			Model 2			Model 3		
		ML	GLS	OLS	ML	GLS	OLS	ML	GLS	OLS
ML	Φ_1	10e-8	10e-8	10e-7	10e-8	10e-8	10e-7	10e-8	10e-8	10e-7
	Φ_2	10e-8	10e-8	10e-7	10e-8	10e-8	10e-7	10e-8	10e-8	10e-7
	Φ_3	10e-8	10e-8	10e-8	10e-8	10e-8	10e-8	10e-8	10e-8	10e-8
	Φ_4	10e-8	10e-8	10e-8	10e-8	10e-8	10e-8	10e-8	10e-8	10e-8
	Φ_5	10e-8	10e-8	10e-8	10e-8	10e-8	10e-8	10e-8	10e-8	10e-8
	Φ_6	10e-8	10e-8	10e-8	10e-8	10e-8	10e-8	10e-8	10e-8	10e-8
Lasso	Φ_1	0.005735	0.005651	0.032370	0.005731	0.005655	0.030312	0.005731	0.005655	0.030312
	Φ_2	0.005916	0.005827	0.027573	0.005994	0.005908	0.027586	0.005994	0.005908	0.027586
	Φ_3	0.012336	0.012098	0.067576	0.013412	0.013133	0.069292	0.014078	0.013769	0.073323
	Φ_4	0.014124	0.013831	0.064840	0.014278	0.013986	0.064981	0.014278	0.013986	0.064981
	Φ_5	0.018302	0.017975	0.082475	0.020346	0.019956	0.086345	0.022978	0.022423	0.096522
	Φ_6	0.026215	0.025463	0.088333	0.027231	0.026465	0.093916	0.027231	0.026465	0.093916

Table 6.25: DA Table - SEM, No Model Error

DA : Discrepancy due to Approximation (FA)

Estimation Method	Φ	Model 1			Model 2		
		ML	GLS	OLS	ML	GLS	OLS
ML	Φ_1	**0.054278**	0.051449	0.010321	0.039953	0.039179	0.004390
	Φ_2	**0.054278**	0.053163	0.022657	0.035946	0.035352	0.007369
	Φ_3	**0.054278**	0.051442	0.022993	0.021352	0.020814	0.006485
	Φ_4	**0.054278**	0.053548	0.046053	0.018059	0.017792	0.012438
	Φ_5	**0.054278**	0.051286	0.031697	0.019812	0.018745	0.008191
	Φ_6	**0.054278**	0.052707	0.056290	0.014266	0.014015	0.011885
Lasso	Φ_1	0.059421	0.055432	0.018369	0.047830	0.047129	0.012715
	Φ_2	0.058481	0.055981	0.028633	0.042892	0.041970	0.015804
	Φ_3	0.089454	0.079313	0.096956	0.058302	0.055102	0.052903
	Φ_4	0.066244	0.065210	0.091665	0.033367	0.032993	0.048666
	Φ_5	0.146680	0.132632	0.230425	0.153171	0.151870	0.296197
	Φ_6	0.088646	0.088858	0.151327	0.054177	0.054930	0.097009

Estimation Method	Φ	Model 3			Model 4		
		ML	GLS	OLS	ML	GLS	OLS
ML	Φ_1	0.039953	0.039179	0.004390	**0.054278**	0.051449	0.010321
	Φ_2	0.035946	0.035352	0.007369	**0.054278**	0.053163	0.022657
	Φ_3	0.021352	0.020814	0.006485	**0.054278**	0.051442	0.022993
	Φ_4	0.018059	0.017792	0.012438	**0.054278**	0.053548	0.046053
	Φ_5	0.019812	0.018745	0.008191	**0.054278**	0.051286	0.031697
	Φ_6	0.014266	0.014015	0.011885	**0.054278**	0.052707	0.056290
Lasso	Φ_1	0.047830	0.047129	0.012715	0.059312	0.055459	0.018540
	Φ_2	0.042892	0.041970	0.015804	0.058491	0.055915	0.028572
	Φ_3	0.058302	0.055102	0.052903	0.089398	0.079254	0.096857
	Φ_4	0.033367	0.032993	0.048666	0.066430	0.065583	0.091883
	Φ_5	0.153171	0.151870	0.296197	0.146643	0.132588	0.230348
	Φ_6	0.054177	0.054930	0.097009	0.088622	0.088826	0.151288

Table 6.26: DA Table - FA, Model Error Involved

DA : Discrepancy due to Approximation (SEM)

Estimation Method	Φ	Model 1			Model 2			Model 3		
		ML	GLS	OLS	ML	GLS	OLS	ML	GLS	OLS
ML	Φ_1	**0.052478**	0.053568	0.049895	0.051074	0.051822	0.050153	0.034204	0.033439	0.014383
	Φ_2	**0.052478**	0.054062	0.083775	0.048996	0.049997	0.083030	0.031467	0.030743	0.017524
	Φ_3	**0.052478**	0.054626	0.061843	0.049836	0.051394	0.062409	0.021031	0.021048	0.014282
	Φ_4	**0.052478**	0.054522	0.099623	0.046413	0.047618	0.100194	0.016315	0.016309	0.013986
	Φ_5	**0.052478**	0.054573	0.067897	0.050161	0.051799	0.067910	0.016456	0.016515	0.014767
	Φ_6	**0.052478**	0.054241	0.095059	0.047467	0.048486	0.095524	0.010941	0.010966	0.010931
Lasso	Φ_1	0.056074	0.056208	0.061399	0.054845	0.054701	0.065602	0.039433	0.038565	0.051580
	Φ_2	0.056091	0.056421	0.088776	0.052594	0.052623	0.091367	0.036294	0.035640	0.047931
	Φ_3	0.062526	0.061888	0.134129	0.060457	0.059488	0.137708	0.035351	0.034988	0.105389
	Φ_4	0.063900	0.062112	0.177476	0.057622	0.056180	0.175576	0.032355	0.031850	0.099542
	Φ_5	0.063537	0.063730	0.145177	0.062635	0.062338	0.157630	0.041533	0.040999	0.140938
	Φ_6	0.075182	0.071032	0.220961	0.070004	0.066444	0.213594	0.044307	0.042861	0.141115

Table 6.27: DA Table - SEM, Model Error Involved

In conclusion, it seems the present L_1-regularized SEM algorithm fails to control the degree of DA, which incurs excessive rise in OD values. In other words, we can expect the Lasso SEM to work better by improving its capability of balancing DA increase and DE reduction.

Another critical point we can observe from the table 6.26 and 6.27 is changes in DA values depending on model specifications. Comparing correctly specified model 1 and those models with misspecification, we can see that DA values of ML estimation and ML discrepancy function show decreasing trends as degree of misspecification increases. This is in common of both FA and SEM. It seems that the added unnecessary paths explain E, which is the component that can not be explained even with the quasi-true model, reducing the model error and DA.

This phenomenon contradicts the premise stating that the quasi-true model, model 1 in our simulations, is the best-approximating model of the true data-generating process. Hence, the result implies that there may be a flaw in the current δ-inserting method. As a matter of fact, even though Cudeck and Browne (1992) devise their procedure based on the three conceptualization of data-generating process in Cudeck and Henly(1991), they do not leave any comments on the concept of the best-approximating model. This issue will be discussed again in the final chapter of the present thesis.

2) Bias Squared

The following four tables contain observed bias squared in FA and SEM. Table 6.28 and 6.29 are obtained in no model error case, while remaining two tables are from model error involved case.

Since the model error is not supposed to affect estimated values, there seems to be little difference in ML's bias estimates whether the model error is employed or not. However, in SEM model 3, where the degree of misspecifica-

Bias Squared (FA)

	Φ	Model 1		Model 2		Model 3		Model 4	
		ML	Lasso	ML	Lasso	ML	Lasso	ML	Lasso
unstandadized	Φ_1	10e-8	0.001310	10e-9	0.002507	10e-8	0.003657	10e-8	0.001312
	Φ_2	10e-9	0.001460	10e-9	0.001460	10e-9	0.001460	10e-8	0.001463
	Φ_3	10e-8	0.008851	10e-8	0.022016	10e-8	0.049731	10e-8	0.008831
	Φ_4	10e-9	0.008131	10e-9	0.008131	10e-9	0.008131	10e-8	0.008387
	Φ_5	10e-8	0.032448	10e-8	0.069710	10e-8	0.175415	10e-8	0.032413
	Φ_6	10e-8	0.037447	10e-8	0.037447	10e-8	0.037447	10e-8	0.037394
standardized	Φ_1	10e-16	0.000584	10e-15	0.001524	10e-15	0.002290	10e-16	0.000586
	Φ_2	10e-17	0.000355	10e-15	0.000355	10e-15	0.000355	10e-16	0.000358
	Φ_3	10e-14	0.019934	10e-15	0.029960	10e-14	0.049754	10e-16	0.019919
	Φ_4	10e-14	0.007288	10e-15	0.007288	10e-16	0.007288	10e-14	0.007411
	Φ_5	10e-14	0.122426	10e-15	0.154584	10e-13	0.244400	10e-14	0.122399
	Φ_6	10e-14	0.044317	10e-14	0.044317	10e-14	0.044317	10e-14	0.044292

Table 6.28: Bias Squared Table - FA, No Model Error

DA : Discrepancy due to Approximation (SEM)

Estimation Method	Φ	Model 1			Model 2			Model 3		
		ML	GLS	OLS	ML	GLS	OLS	ML	GLS	OLS
ML	Φ_1	10e-8	10e-8	10e-7	10e-8	10e-8	10e-7	10e-8	10e-8	10e-7
	Φ_2	10e-8	10e-8	10e-7	10e-8	10e-8	10e-7	10e-8	10e-8	10e-7
	Φ_3	10e-8	10e-8	10e-8	10e-8	10e-8	10e-8	10e-8	10e-8	10e-8
	Φ_4	10e-8	10e-8	10e-8	10e-8	10e-8	10e-8	10e-8	10e-8	10e-8
	Φ_5	10e-8	10e-8	10e-8	10e-8	10e-8	10e-8	10e-8	10e-8	10e-8
	Φ_6	10e-8	10e-8	10e-8	10e-8	10e-8	10e-8	10e-8	10e-8	10e-8
Lasso	Φ_1	0.005735	0.005651	0.032370	0.005731	0.005655	0.030312	0.005731	0.005655	0.030312
	Φ_2	0.005916	0.005827	0.027573	0.005994	0.005908	0.027586	0.005994	0.005908	0.027586
	Φ_3	0.012336	0.012098	0.067576	0.013412	0.013133	0.069292	0.014078	0.013769	0.073323
	Φ_4	0.014124	0.013831	0.064840	0.014278	0.013986	0.064981	0.014278	0.013986	0.064981
	Φ_5	0.018302	0.017975	0.082475	0.020346	0.019956	0.086345	0.022978	0.022423	0.096522
	Φ_6	0.026215	0.025463	0.088333	0.027231	0.026465	0.093916	0.027231	0.026465	0.093916

Table 6.29: Bias Squared Table - SEM, No Model Error

tion is not small, there are some changes in ML result. Maybe the model error has influence on ML in combination with misspecification, which is equal to model difference discussed in the last chapter.

Bias Squared (FA)

	Φ	Model 1		Model 2		Model 3		Model 4	
		ML	Lasso	ML	Lasso	ML	Lasso	ML	Lasso
unstandadized	Φ_1	10e-8	0.001675	0.007414	0.016381	0.029705	0.044791	10e-8	0.001581
	Φ_2	10e-9	0.001670	0.003379	0.004154	0.012276	0.009469	10e-8	0.001730
	Φ_3	10e-8	0.017490	0.019915	0.088284	0.216815	0.373903	10e-8	0.017439
	Φ_4	10e-9	0.011187	0.006950	0.017886	0.07787	0.062574	10e-8	0.011472
	Φ_5	10e-8	0.146634	0.031701	2.079894	0.53036	3.183969	10e-8	0.146503
	Φ_6	10e-8	0.083056	0.011310	0.091507	0.202894	0.200009	10e-8	0.082952
standardized	Φ_1	10e-13	0.000575	0.006038	0.012947	0.024421	0.036101	10e-13	0.000560
	Φ_2	10e-13	0.000329	0.002766	0.002456	0.010214	0.006634	10e-14	0.000354
	Φ_3	10e-14	0.028849	0.011975	0.080867	0.137408	0.286009	10e-13	0.028800
	Φ_4	10e-13	0.011163	0.004581	0.017500	0.059118	0.056138	10e-14	0.011341
	Φ_5	10e-12	0.214363	0.012727	0.948919	0.251942	1.287277	10e-12	0.214273
	Φ_6	10e-13	0.086707	0.005375	0.096212	0.134535	0.192045	10e-12	0.086647

Table 6.30: Bias Squared Table - FA, Model Error Involved

Bias Squared (SEM)

	Φ	Model 1		Model 2		Model 3	
		ML	Lasso	ML	Lasso	ML	Lasso
unstandadized	Φ_1	10e-8	0.005540	10e-8	0.006492	0.005254	0.012179
	Φ_2	10e-8	0.005619	10e-8	0.006585	0.011256	0.017270
	Φ_3	10e-8	0.014357	10e-8	0.018711	0.018326	0.040060
	Φ_4	10e-8	0.017312	10e-8	0.020247	0.027236	0.046875
	Φ_5	10e-8	0.028425	10e-8	0.043411	0.037431	0.074709
	Φ_6	10e-8	0.055790	10e-8	0.061051	0.033838	0.091637
standardized	Φ_1	10e-12	0.000776	10e-15	0.002389	0.004831	0.006368
	Φ_2	10e-12	0.000966	10e-14	0.002791	0.011440	0.012732
	Φ_3	10e-11	0.008149	10e-14	0.013347	0.015445	0.029187
	Φ_4	10e-12	0.009457	10e-14	0.014199	0.025763	0.037408
	Φ_5	10e-11	0.019501	10e-14	0.039154	0.026178	0.053635
	Φ_6	10e-12	0.044285	10e-13	0.050829	0.027657	0.074713

Table 6.31: Bias Squared Table - SEM, Model Error Involved

6.3.2 Correlation Analysis of Fit Indices

Note that one of our research questions is to figure out which of the goodness of fit indices can be used as indicators of generalizability and reproducibility. For this purpose, we analyze correlations between generalizability criteria such as OD and sMSE, and several fit indices generally used in SEM. Each of these indices was introduced in Chapter 1 and 5. Furthermore, we also investigate the correlational relationships among discrepancy indices. This is to study about an implicit assumption that even though SEM deals with only the sample discrepancy, the result may not contradict the pursuit of reproducible outcomes. Also connection between discrepancies in SEM and parameter discrepancy, which indicates MSE, bias squared, and variance, can be learned by this analysis.

A correlation analysis is somewhat naive to reveal the underlying relationships among the indices we investigated; In fact, careful scrutiny with regard to their mathematical properties should be carried out in order to satisfy our purpose. However, this study surely helps us draw a big picture of the issue.

Considering all fit indices introduced in the present thesis, A 15×15 correlation table is computed(13 fit indices, and OD, sMSE). However we only post a part of the table since our purpose is to study on the generalizability and reproducibility. For this purpose, it is sufficient to present fit indices' correlation with OD and sMSE. Furthermore, note that we can obtain a number of correlation tables according to combinations of 1) types of discrepancy function and 2) estimation methods exploited to obtain results. Due to the lack of space, results from only $2 \times 2 = 4$ cases are presented; OLS and ML for discrepancy functions, ML and Lasso for estimation methods.

Also, we can obtain this correlation table for each of conditions and models.

This yields $4 \times 6 \times 7 \times 2 = 336$ tables; 4 for sample sizes, 6 for Φ matrices, 7 for FA and SEM models, and 2 for the model error condition. It is impractical to present all of them. Therefore, we omit some outcomes and abbreviate the remains as follows. First, we select SEM model 3 with model error as the most comprehensive one; general SEM model with several misspecified parameters and perturbations in true process. Regarding $4 \times 6 = 24$ conditions, we obtain an average of 24 correlation tables. Obviously this approach may lose some informations, but fundamental relationships among fit indices can be summarized efficiently in this way. Standard deviations of each of correlations are $0.07 \sim 0.14$. Thus readers should be cautious interpreting the results that the values may vary over conditions even though the degree is not that large.

The outcome is contained in table 6.32. In general, fit indices show poor relationships with generalizability indices. As shown in other tables below, this seems due to the fact that SD has very small correlations with OD and sMSE.

	Discrepancy Function : F_{OLS}				Discrepancy Function : F_{ML}			
	Method : Lasso		Method : ML		Method : Lasso		Method : ML	
	OD	sMSE	OD	sMSE	OD	sMSE	OD	sMSE
OD	1.000	.299	1.000	.141	1.000	.376	1.000	.307
sMSE	.299	1.000	.141	1.000	.376	1.000	.307	1.000
RMSEA	.323	.049	.004	.037	.070	.034	.010	.021
NFI	-.324	-.059	.004	.037	.006	-.043	-.026	-.048
TLI	-.325	-.056	-.049	-.063	-.053	-.044	-.008	-.024
CFI	-.330	-.056	.032	.035	-.053	-.042	-.013	-.033
RFI	-.323	-.061	.064	.075	.008	-.052	-.026	-.048
IFI	-.326	-.056	.029	.037	-.055	-.043	-.008	-.024
RNI	-.326	-.056	-	-	-.054	-.043	-.008	-.024
$\hat{\gamma}$	-.336	-.056	-.024	-.059	-.074	-.041	-.007	-.020
MC	-.335	-.054	.019	-.016	-.074	-.040	-.007	-.019
SRMR	.356	.046	-	-	.184	.046	.035	.065
ECVI	.337	.060	-.024	-.059	.073	.048	.007	.020
AIC	.138	.016	.019	-.016	-.097	.016	-.312	-.035
BIC	.138	.020	-.004	-.038	-.098	.020	-.312	-.035

Table 6.32: Correlations Between Fit Indices and Generalizability Indices

Since the most of fit indices are defined based on SD, they are not able to be any indications of reproducible results. A remarkable point can be found in comparing the outcomes from each of combinations. In case that F_{ML} is exploited, the outcome is desperate. Most correlations with OD and sMSE are close to zero, especially when we use ML method to fit the model. Lasso result contains higher values, but in only a slight degree. The outcome does not change when we obtain correlations by ML method with F_{ML}. However, some noteworthy change is observed with the (Lasso, F_{OLS}) combination. Even though the outcome is not much satisfactory, many of correlations of fit indices are increased to higher than 0.3. Those values per se may not be significant and meaningful. However, it seems that Lasso is able to retrieve the relationship between generalizability indices and general model fit indices in SEM.

Against the expectation, information criteria and ECVI, which are considered as indices indicating the generalizable outcome, do not produce any significant results. Things get even worse with (F_{ML}, ML) condition where AIC and BIC yield negative correlations, which is supposed to be positive, with OD and sMSE. Lasso with F_{OLS} produces the better, but not great results. however, note that this phenomenon should have limited interpretation since they are valuable only when we compare various models with these criteria. Absolute magnitude of those indices, and thus their correlation with other indices, may not be appropriate for deriving results with meaningful interpretation for generalizability and reproducibility.

Another thing we should point out is the correlation between OD and sMSE. As other fit indices, their relationship can be retrieved by Lasso estimation. One difference occurs in that for OD and sMSE, F_{ML} produce higher correlations than F_{OLS}.

Next, correlations among discrepancy criteria are presented in table 6.33 and

6.34. Those tables are obtained in the same way we described above for table 6.32. Table 6.33 contains correlations based on F_{OLS}, while F_{ML} is used for table 6.34. In both tables, low triangular part represents the Lasso result and the ML result is shown in upper triangular with gray-colored cells for distinction.

Interpretations for those tables can be abbreviated with the following list.

1. ML estimation is able to produce less DA than Lasso. Therefore, DE shows high correlation with OD by this method. Since the model error is employed regarding the ML discrepancy function, ML yields the coefficients little short of 1 with F_{ML}.
2. However, correlations between SD and OD are no better than zero when we exploit the ML estimation. It seems that this is due to the fact that, minimizing SD, ML captures even the inherent features specific only to the current sample. Misspecified parameters can be of help reducing this kind of variability in the sample, which yields non-generalizable outcomes.
3. In the context of 2), Lasso seems to be able to remedy the problem with its complete shrinkage ability. This point can be seen by the fact that the correlations between SD and OD are recovered in a quite noticeable degree, especially when F_{OLS} is exploited.
4. Lasso also exerts its advantage in improving correlations between matrix discrepancies and parameter discrepancies. By removing unnecessary parameters, parameters can have more intimate relations with their covariance matrix.
5. But the correlations between two types of discrepancies are higher in F_{ML} results than F_{OLS}.
6. SD.cv, our cross-validation criterion for Lasso, yields high correlations

F_{OLS}	OD	DE	SD.cv	SD	sMSE	MSE	var
OD	-	.988	.849	***.004***	.141	.149	.246
DE	.740	-	.837	.001	.132	.139	.246
SD.cv	.821	.607	-	.005	.110	.118	.211
SD	***.336***	.159	.267	-	.037	.039	.023
sMSE	.299	.301	.227	.056	-	.917	.431
MSE	.391	.357	.297	.100	.871	-	.610
Var	.399	.539	.314	.055	.434	.669	-

Table 6.33: Correlations Among Matrix/Parameter Discrepancies - F_{OLS} case

which are stronger than 0.8 over all conditions.

7. sMSE and MSE show strong relationships; the coefficient is higher than 0.85, 0.90 for Lasso and ML respectively.

8. Variance produces lower correlation for sMSE than MSE. This should be the case since the variance is computed with unstandardized estimates. If we obtain variances of standardized coefficients, the opposite is expected.

F_{ML}	OD	DE	SD.cv	SD	sMSE	MSE	var
OD	-	.999	.815	***.007***	.307	.383	.603
DE	.888	-	.816	.009	.312	.388	.610
SD.cv	.846	.751	-	.008	.259	.319	.512
SD	***.076***	.028	.067	-	.020	.026	.037
sMSE	.376	.371	.321	.034	-	.917	.431
MSE	.492	.460	.414	.057	.871	-	.610
Var	.609	.687	.512	.042	.434	.669	-

Table 6.34: Correlations Among Matrix/Parameter Discrepancies - F_{ML} case

In the investigation of the relationship between SD and OD, an additional information is provided in table 6.35. As we described, tables above are based on the averages of correlations over the 24 conditions regarding sample sizes and Φ matrices. This averaging can drop some informations, some of which have meaningful implications. In this regard, we add the table containing correlation coefficients of SD and OD, computed for each of the conditions.

From the result, we obtain several additional implications. Af first, correlations between SD and OD vary severely when F_{ML} is used in evaluating dis-

Sample Size	Φ	Discrepancy Function : F_{OLS}		Discrepancy Function : F_{ML}	
		Lasso	ML	Lasso	ML
150	Φ_1	.155	.063	.107	.024
	Φ_2	.267	-.096	.003	-.048
	Φ_3	.331	-.089	.134	.118
	Φ_4	.180	-.101	.061	.084
	Φ_5	.172	-.189	-.212	-.205
	Φ_6	.216	.128	.041	.049
250	Φ_1	.284	-.047	-.096	-.076
	Φ_2	.266	.052	.036	-.028
	Φ_3	.512	-.101	.058	.056
	Φ_4	.293	.117	.007	.038
	Φ_5	.218	.066	.041	.078
	Φ_6	.346	.080	.016	-.021
500	Φ_1	.529	.014	.177	.114
	Φ_2	.413	-.150	.083	-.002
	Φ_3	.499	.183	.100	.037
	Φ_4	.400	-.003	-.031	-.139
	Φ_5	.402	-.068	.130	.068
	Φ_6	.237	.074	.117	.018
1000	Φ_1	.588	.046	.272	.130
	Φ_2	.455	.139	.009	-.053
	Φ_3	.356	-.182	.112	.102
	Φ_4	.223	-.001	.007	-.170
	Φ_5	.235	.037	.160	-.090
	Φ_6	.491	.134	.490	.073

Table 6.35: Correlations Between OD and SD Computed for Each of Conditions

crepancies. Even signs of coefficients change depending on conditions. This can yield severe problems when we use the SEM results obtained by ML estimation; Judging based on OD as a generalizability index, what we expect from the outcome may not be applicable to other independent samples and the real world phenomenon we are interested in.

ML result still has the same trend even with F_{OLS}. However, Lasso produces a remarkable outcomes when the discrepancies are measured by the OLS function. Correlation coefficients have stable signs over all the conditions only in this combination(Lasso, F_{OLS}). Also, coefficients show fairly strong magnitudes. This outcomes support the argument that L_1-regularization can improve SEM on producing more generalizable and reproducible results.

Chapter 7

Discussion

In the present thesis, the author tried to implement Lasso-type Regularization to SEM and yield more generalizable, reproducible SEM results. For this purpose, we discussed the overall discrepancy and mean squared error as indices representing generalizability and reproducibility.

In addition, we studied Bayesian Lasso SEM, which is a preceding attempt to regularize SEM with L_1 penalty, and pointed out its fundamental limitation; This Bayesian approach cannot produce zero coefficients and delete misspecified parameters but only yields posterior distributions with close-zero central tendencies.

Furthermore, we suggested the method to fit Lasso SEM, as a better alternative of BLasso SEM. After developing Lasso CFA by modifying Lasso EFA in former researches, we expanded the method to L_1-regularized structural model. In order to estimate the model, regularized conditional expectations of log-likelihood are derived for both of measurement and structural model. Finally, 'Double EM-Algorithm' was proposed in order to optimize both likelihoods simultaneously.

In pursuit of testing performance of the proposed method, intense simulation studies were carried out considering various conditions such as model er-

ror, degree of misspecification, magnitudes of covariance matrix, sample sizes, and so on. Moreover, overall discrepancy, mean squared error, sample discrepancy and other widely used fit indices are observed and compared for studying their relationships and investigating our research questions.

Limitations

Even though the author attempted to investigate relations among SEM, regularization, and generalizability/reproducibility as thoroughly as possible, some limitations need to be pointed out.

First of all, the simulation result did not always support the proposed method. Lasso outperformed and fell behind the existing method by conditions and criteria. For accurate and broad application of Lasso, studies should be carried out on the conditions in which Lasso is able to produce better outcomes. Also, remedies to improve Lasso's performance, especially in conditions that the method didn't keep up with ML, have to be inspected.

In the similar context, secondly, there might be several drawbacks intrinsic in the proposed algorithm. In particular, it seems that the optimization method used in the present simulation should be modified in order to improve Lasso's complete shrinkage ability. This point will be discussed below. Another issue is also about the optimization method of Lasso SEM. The present method approached the problem in a slightly different way from the conventional method. That is, we exploited conditional independence of each equation in SEM to implement L_1-regularization, not minimizing the discrepancy measured by F function. In order to study another possibility, it will be worthwhile to apply regularization to the original objective function of SEM. More specifically, an optimization based on minimizing '$F(S, \hat{\Sigma}) + \kappa \times$ penalty' de-

serves our attention.

Thirdly, even though we took into consideration comparatively comprehensive conditions, including degree of misspecification, sample sizes, and different Φ matrices, there might be some other conditions which have to be covered. For example, in the Research 2, where we analyzed properties of Lasso SEM, manipulation on Φ matrix was not able to cover more various conditions of Σ_ω. Despite the research results implying that conditions we regarded on Φ brought about different performances of Lasso regularization, other possible conditions on Σ_ω should have been considered. Manipulation on B and Γ matrices with Φ is worth to be investigated regarding this issue, even though combinations of conditions on those matrices may yield a huge number of cases. Additionally, the models studied in the present simulation studies are not able to embrace all the cases. The more general form of the structural equation model, and its variations should be studied with varying degree of misspecifications.

Last but not the least, lack of real data analysis may be another limitation as well. Since the true or quasi-true model inherent to practical sample is unknown and cannot be known, performance of Lasso is not able to be evaluated in this way. The present thesis focused on producing simulation outcomes which enable us to investigate data-generating process and several unobservable indices such as OD and MSE. However, it will yield some meaningful results to apply Lasso SEM to models and data previously analyzed by maximum likelihood estimation. The models which have satisfactory goodness of fit indices but contain insignificant estimates will attract huge attention.

Implications

In spite of some limitations mentioned above, the present study has some significance as well. Above all, it is the first study that attempts to implement Lasso to the structural equation modeling. Unlike the existing estimation methods in SEM, which produce the result maintaining the model inputted by its user, the present thesis tried to obtain parsimonious and sparse models by detecting and removing unnecessary parameters during estimation procedure by means of L_1-regularization.

This attempt can make a contribution for a paradigm-shift that frees SEM from the convention excessively focusing on model fit indices. It also leads us to interesting issues such as specific forms of models, individual paths, and each of parameters in the structural equation modeling. For instance, sometimes SEM results are published with excellent model fit indices but at the same time with some insignificant estimates. Generally, implications in this insignificance usually are not discussed fully and seriously; Anyway, it has an excellent model fit! Application of Lasso can be regarded as a valuable attempt to modify this kind of analysis result and to help an accurate and deep understanding of underlying phenomenon.

The present thesis studied on BLasso, the preceding attempt to regularize SEM by Bayesian approach, and pointed out its fundamental limitations regarding the complete shrinkage property. This estimation method was also included in the simulation studies and compared with ML principle.

The Monte Carlo experiments mainly dealt with comparison of Lasso and ML results. The simulation was carried out considering various conditions including sample sizes, and several parameter values. This is distinguishable from the preceding researches which considered only one fixed value per each parameter. On manipulating parameters, $3 \times 2 = 6$ types of Φ matrices were

exploited. This may seems like not to manipulate other parameters. However, this is an efficient approach to manipulate the covariance structure implied in population covariance matrix. In fact, if we manipulate Φ and another matrix simultaneously, its effect can be offset. To be specific, with fixed Ψ, it can be the case that $\Lambda_1\Phi_1\Lambda_1^T = \Lambda_2\Phi_2\Lambda_2^T$ even though $\Lambda_1 \neq \Lambda_2$ and $\Phi_1 \neq \Phi_2$. Therefore, out of Λ and Φ, we should fix one and manipulate the other to cover various conditions on population covariance matrix. The present thesis chose to fix Λ and manipulate Φ.

Furthermore, the simulation studies encompassed the process of generating population, which enabled us to observe and investigate several discrepancy concepts such as overall discrepancy, discrepancy due to approximation, and discrepancy due to estimation. These concepts has been studied only in a conceptual way. Our research results provided empirical informations on those indices, particularly their relationships with practically observable fit indices. What is more important is that the research included the investigation on the parameter discrepancies such as mean squared error, bias, and variance of estimators. They has not been dealt with properly before, despite having meaningful informations as matrix discrepancies.

Research results we obtained from simulation studies can be refined by answering the following research questions asked in Chapter 5.

1) Do the Lasso's capability of shrinkage estimation and variable deletion also appear in SEM? That is, is it possible to obtain more parsimonious and sparse SEM results by Lasso regularization?

- For correctly specified parameters, Lasso SEM did not always produce estimates smaller than ML estimates. In regression models, the OLS and ML estimators have a closed form solution. And regularized estimates are always less

than solutions to the normal equation. However, SEM does not have closed form solution for both OLS and ML estimation. Furthermore, the Lasso SEM algorithm proposed in the present thesis exploits a totally different procedure from that of ML estimation. Therefore, shrinkage in comparison with the original ML estimates does not always occur. In fact, the proposed method shrinks the OLS estimates of each individual equation, obtained using the conditional independence.

However, regression coefficients of the structural model produced less mean values, and their distributions located at the left side comparing to ML results. That is, shrinkage effect appeared evidently. Factor loadings in the measurement model did not show the same tendency.

For misspecified parameters, there is a question mark over Lasso's complete shrinkage ability. Obviously, we observed that Lasso SEM was able to produce zero-valued coefficients, which means the method detected and removed some unnecessary parameters from the model. However, the performance of Lasso varied according to whether the model error was employed to population-generating process or not. When no model error was assumed, Lasso yielded great results. On the contrary, in case the model error is involved, the proportion of complete shrinkage declined by a third.

Differences between two cases became more evident as sample sizes increased. Note that, the typical unbiased sample covariance matrix S has consistency. Thus, in a large-scale sample, the matrix is asymptotically equal to the population covariance matrix Σ_0. In this regard, when n is large enough, S reflects well the true structure implied in Σ_0.

If there is no or less model error, this structure consists mostly of covariation among the interrelated variables, not misspecified ones. Thus, Lasso just has to screen out only the sampling error. Especially when the sample size is large,

the sample covariance matrix contains pure systematic relationships supposed to be generated by the operating model. It seems that the current Lasso SEM has a capability of capturing the system and deleting superfluous coefficients.

When the degree of model error is non-ignorable, Lasso should overcome both model error and sampling error. Note that the model error does not vanish from the sample covariance matrix even when n is large enough. Therefore, it can be misinterpreted as if the model error represents some meaningful structure in population-generating process. This relation, caused by model error, can affect variabilities of misspecified parameters, which makes Lasso SEM mistake them as core components of true data-generating process. In consequence, the proposed method may not get rid of those unnecessary parameters. Additionally, the goal becomes more difficult due to the fact that the sampling error can be amplified by the model error.

However, this problem regarding the model error, should be limited to the current Lasso SEM algorithm and its optimization method. The point is that optimization of Lasso was carried out based on the minimization of OD or SD.cv in the present thesis. If these criteria could not represent the degree to which the estimation method retrieves variables and paths in the best approximating model, the current optimization strategy for Lasso SEM cannot show the best complete shrinkage ability. Therefore, there might be another approach to optimize Lasso SEM producing more zero-estimates. These issues on OD as an optimization criterion, and on alternative methods of optimizing Lasso SEM will be discussed further below.

2) Can the Lasso reduce the generalizability/reproducibility indices such as OD or MSE when it is applied to SEM?

- We observed that Lasso produced less values of OD and sMSE than those of

ML in several conditions. Therefore, in practical data analysis, researchers can take advantage of Lasso SEM in order to produce more generalizable and reproducible results if the sample covariance matrix is close to any of conditions in the present simulations.

Tables containing DE values in Appendix A indicate that Lasso SEM is capable of reducing DE in a considerable degree. This is the essence of regularization. However, DA tables in subsection 6.3.1 tell us that the proposed method suffers from larger DA than ML. Also the difference was fairly large. This fact implies that some modification should be carried out in order to redress the balance between large DA and small DE, as the bias-variance trade-off in MSE analysis. And this can improve Lasso SEM's performance on producing generalizable and reproducible analysis results.

Another important issue is related to the consistency of the trends in comparison of ML and Lasso SEM. The most dominant condition was magnitude of covariance matrix. Lasso was able to reduce OD in Φ_1 and Φ_2 conditions, where the matrices yielded the highest values covariance matrix out of all the conditions. For Φ_3 and Φ_4, where the moderate covariance matrices were produced, Lasso also show better results than ML as long as the sample is too large. At last in conditions of Φ_5 or Φ_6, ML overwhelms Lasso in reducing OD values. This tendency did not vary largely by other conditions.

Against our expectation, the degrees of misspecification did not exert a strong influence on the OD results. This may arouse criticism that misspecification conditions of the present study were not comprehensive enough to draw meaningful differences. However, the author is of opinion that the matrix discrepancies are not sensitive to parameter misspecification. This point will be discussed more in question 4) below.

In contrast to OD, sMSE showed a remarkable difference depending on de-

grees of misspecification. Lasso is able to reduce this index when some unnecessary parameters are included in model specification. This trend appeared evidently for Φ_1 and Φ_2 conditions. It seems that this is due to Lasso's capability of shrinking those parameters, and the fact that ML cannot distinguish the misspecified parameters form the correctly specified ones.

3) Do BLasso estimates not shrink completely to zero? How about considering them to be zero if their values are less than a pre-determined bound?

- In the present simulation results, there was no case that BLasso yielded exact-zero estimates. The strategy using pre-determined bound can be applied to both of BLasso and ML, which could not produce any remarkable differences between two methods. Also when the bound was smaller than 10^{-2}, the proportion of the case that the estimate can be judged to be removed was no better than zero. Therefore, it is hard to say that BLasso SEM was able to yield zero estimates and delete corresponding parameters.

4) Do overall discrepancy and mean squared error have close relationships empirically?

- We found that parameter discrepancies acted in a slightly different way from matrix discrepancies, though they are conceptually almost the same. For example, while sMSE was affected severely by varying degree of misspecification, OD was not. Also, it turned out that they are not highly correlated. Therefore, it can be concluded that those matrix-based criteria may have some disjunction with the parameter-based indices.

Additionally, in a different point of view, assume that we add a couple of paths to an existing SEM model. In this case, two models - original and modified, are completely different models; have different forms, different numbers

of parameters, different degrees of freedom, and so on. However, their implied covariance matrices may show only a slight difference in their magnitudes as long as the added paths have dominant influences. In this regard, those two models produce a fairly large difference in parameter discrepancies, but only a small difference in matrix discrepancies.

Hence, despite the similarity in their definitions and meanings, those two types of indices have different implications on generalizability and reproducibility of SEM results. This conclusion may also provide some clues for improving Lasso SEM algorithm.

5) Is the existing optimization method of SEM, which is based on minimizing the sample discrepancy, able to produce a generalizable outcome? That is, does the sample discrepancy show high correlation with the overall discrepancy?

- In the correlation analysis, SD showed near-zero correlation with OD and sMSE in ML results. Therefore, it is hardly justifiable to generalize this results.

In fact, this problem seems to be caused by the limitation of ML estimation for SEM that the method produce a result maintaining all the paths and variables in a given model. Note that the higher the model complexity, the more SD can be reduced. This is due to the fact that even superfluous parameters can capture the residual variability which cannot be explained by the true systematic relationships. This residual variance includes inherent variation specific only to the current sample. Therefore when, the model with high degree of misspecification is given to SEM program, ML may diminish SD by taking advantages of those unnecessary parameters. However, it cannot produce the generalizable result reflecting the real world phenomena. Low correlation of SD and OD in ML principle may be attributable to this regard.

One remarkable point is that in analysis of Lasso SEM with F_{OLS}, correla-

tions among OD, and SD and fit indices were recovered in large degree. Especially, correlations between OD and SD, which were obtained for each of conditions, showed stable signs and considerable magnitudes. This improvement implies that Lasso is able to produce more generalizable results than its counterpart, ML estimation. However, correlations between sMSE and OD were low in all the conditions, even when the Lasso is applied with the OLS discrepancy function.

6) Among the model fit indices generally used in SEM, which of them can play a role as an indicator of generalizability and reproducibility? That is, which of model fit indices have empirically meaningful relationships with OD or MSE?

- According to the result in the correlation analysis, fit indices didn't show high correlations with OD and sMSE. This is a reasonable consequence since most of them are defined based on SD.

Again, we concluded that the original estimation method for SEM is not able to produce generalizable and reproducible results.

In addition to the above, we shall discuss the following further issues.

i) Are OD and sMSE appropriate to be criteria indicating generalizable and reproducible results?

- Obviously, by their definitions. However, note that the generalizability and reproducibility are very broad and comprehensive notions. Therefore those two indices cannot cover all the aspects of reproducible results. In fact, it seems that each of them reflects different features of generalizability, yielding low correlations between OD and sMSE over all the conditions.

Lasso seems to be able to improve some aspects of generalizability; the correlation between OD and SMSE was recovered by Lasso when those indices

are evaluated with F_{OLS}. Advantages of Lasso, including its complete shrinkage ability, can be revealed more by other criteria and notions.

ii) Is there any other approach to improve Lasso's complete shrinkage ability?

- A difference between OD and sMSE provides some implications for this question. Note that the proposed Lasso SEM estimation is based on minimizing the F-value of OD or SD.cv. In this process, an optimization is carried out by means of minimizing differences between input covariance matrices. In other words, it focuses on recovering covariational relationships implicit in the target covariance matrix, rather than finding significant paths, removing misspecified parameters, and reconstructing the best-approximating data-generating process. Also, the same comment can be applied to ML and other existing methods since their whole optimization procedure depends on F value.

With this features, misspecified and unnecessary parameters can be used in an effective way by letting them capture and reduce the model error and the sampling error in the optimization process of this kind. In this manner, those parameters are not removed from the model, yielding a smaller minimum value of OD but the best-approximating model cannot be achieved.

Therefore, with some modification on Lasso estimation, there is a possibility that the L_1-regularization yields the best shrinkage estimation even in the case the model error is employed. This modification can be carried out by changing the Lasso tuning criterion. For example, we can exploit the method similar to that in linear regression analysis. In that field, the tuning parameter is determined to minimize a difference between the future response y_0 and its predicted value $\hat{y}_0$. This can be conducted using the expected prediction error $E((y_0 - \hat{y}_0)^2)$ as an optimization criterion(Hastie, Tibshirani, & Friedman,

2008). Generally, a cross-validated estimates replaces the expected prediction error in practice.

The problem is that regressors in the measurement model and both responses and regressors in the structural model are unobservable latent variables in SEM so that the approach in linear regression has not been used. An alternative can be found in the conditional distribution of ω which was derived in Chapter 4. That is, we can estimate ω by its conditional expectation and compute the estimated prediction error. In this way, Lasso SEM optimization may produce a result different from what we observed in the present thesis; larger F-values of OD, but more complete shrinkage of misspecified parameters which yields an improved ability to reconstruct quasi-true structure.

Another important issue can be found in terms of overall discrepancy. In the present thesis, we did not give much distinction between OD and its F value. However, it should be noted that OD can be measured in different ways. Various definitions of the discrepancy function - ML, GLS, and OLS can be a good example. MSE and sMSE are other candidates.

As we described above, F discrepancy functions of OD, which are defined in terms of covariance matrices, are not the best criteria for reproducing accurate paths and variables in a model. Any Replacement of F-value based evaluation of overall discrepancy and generalizability may lead an improvement. The alternative criterion should be defined with care for reflecting each parameters in quasi-true model, not merely its covariance matrix.

iii) Decomposition of model error, and its relation with ML and Lasso

- One more issue will be covered regarding model error, misspecification, and Lasso. In practical research, the only thing we know is our hypothetical model $\Sigma_k(\cdot)$. not the best approximating model $\Sigma_0(\cdot)$. Let E_0 be the model error, the

component of true population-generating process that cannot be explained by $\Sigma_0(\cdot)$. Then, from the conceptualization c) of population covariance matrix, the model error corresponding to $\Sigma_k(\cdot)$ can be expressed in the following manner.

$$\begin{aligned}
\Sigma_0^* &= \Sigma_0(\theta_0) + E_0 \\
\Sigma_0^* &= \Sigma_k(\tilde{\theta}_k) + \left[\Sigma_0(\theta_0) - \Sigma_k(\tilde{\theta}_k)\right] + E_0 \\
\Sigma_0^* &= \Sigma_k(\tilde{\theta}_k) + E_k \\
&\Rightarrow E_k = \left[\Sigma_0(\theta_0) - \Sigma_k(\tilde{\theta}_k)\right] + E_0
\end{aligned}$$

That is, since $\Sigma_k(\cdot)$ is not the quasi-true model, the model error of $\Sigma_k(\cdot)$ consists of 1) difference between $\Sigma_k(\tilde{\theta}_k)$ and $\Sigma_0(\theta_0)$ and 2) the model error of $\Sigma_0(\cdot)$. Also note that 1), which henceforth we call as 'Model Difference', includes the difference between model functions themselves and the difference between true parameter θ_0 and its estimate $\tilde{\theta}_k$. Since $\Sigma_0(\cdot)$ is unknown and cannot-be-known, these two components are not distinguishable in practice. Despite of this fact, discussion on each of these components can provide valuable implications on the relationship among Lasso, misspecification, and model error.

Assume that the whole population is available and used for comparing the estimation of ML and Lasso SEM. The component 2), model error E_0, cannot be reduced by the complete shrinkage of Lasso. This error is supposed to be minimized during the estimation with the quasi-true model $\Sigma_0(\cdot)$ so that yields δ when it is evaluated by an appropriate discrepancy function. Misspecification also cannot affect the error E_0 since they are only related to $\Sigma_k(\cdot)$.

The only changeable component is 1) the model difference. Employing some unnecessary parameters influences this part. When the resulting model is more different from $\Sigma_0(\cdot)$ than the original model does, model difference always increases and yields larger F-values. This is a natural consequence considering the definition of the best approximating model $\Sigma_0(\cdot)$ suggested by Cudeck and

Henly(1991); There is no model with fewer than $p(p+1)/2$ parameters that approximates Σ_0^* better than $\Sigma_0(\theta_0)$. Therefore, all models are unable to yield smaller model error than E_0 and δ, as long as it includes as many parameters as the saturated model – which is completely meaningless in practice.

The greatest strength of Lasso can be exerted regarding this point; Lasso is able to reduce the model difference. When our model $\Sigma_k(\cdot)$ employs more parameters than $\Sigma_0(\cdot)$, Lasso captures and leaves out the misspecification, producing less model difference. However, this argument is valid only when the Lasso estimation on correctly specified parameters is as good as ML's. In other words, Lasso should work well even when there is no misspecification. If this is not the case, Lasso causes the worse DA than ML, for the same amount of model error E_0. In turn, F-value of OD can also be aggravated. In the present study, this problem occurs especially for Φ_5 and Φ_6, where ML almost always outperforms Lasso(check the DA table in Subsection 6.3.1).

Cudeck and Browne procedure, the method to insert model error, is the last thing we should discuss(readers interested in this procedure can refer to Appendix E. As Cudeck and Henly(1991)'s comment says, there should be no model that is able to yield less model error than the best-approximating model $\Sigma_0(\cdot)$. Therefore, the procedure should treat the model used for generating data in simulation as the best approximating model, satisfying those conditions in Cudeck and Browne(1992) as well. However, it seems that the existing procedure does not reflect this prerequisite. Cudeck and Browne(1991) proposed the following conditions for the method which aims to insert model error properly.

i) $\underset{\theta}{\text{argmin}}(F(\Sigma_0^*, \Sigma_0(\theta))) = \theta_0$.
ii) For some predetermined value of δ, $F(\Sigma_0^*, \Sigma_0(\theta_0)) = \delta$.

DA tables in subsection 6.3.1 contain the exact δ value for model 1 when the error is evaluated by F_{ML}. In this case, where there is no misspecification, the given parameter values are recovered perfectly. In other words, the existing δ-insertion method works well for the conditions above.

However, for the other models that contain some degree of misspecification, DA turned out to be decreased. This is exactly the opposite of what theoretically can be expected. It seems that misspecified parameters explain the variability of population which cannot be explained by $\Sigma_0(\cdot)$. However, this variability is from the model error E_0, which is defined as the component that cannot be explained by any systematic models. In this regard, something is in complete contradiction with the definitions of the best-approximating model and model error. This problem causes that ML can reduce the model error by additional misspecification, which is totally inappropriate. Hence, there is a possibility that the whole F-values of OD obtained for ML method is estimated better than what it really is. The comparison study of ML and Lasso SEM in the present thesis should be interpreted with this in mind.

The improvement can be achieved on the model error inclusion procedure by adding the following requisite to the conditions above.

iii) The model used in data-generation is defined to be the best approximating model $\Sigma_0(\cdot)$. That is, other models cannot produce less model error as long as they have more than $p(p+1)/2$ parameters.

Suggestions

We conclude the present thesis by providing some suggestions for the future researches.

Above all, other approaches to regularize SEM with Lasso are worth trying. In the present thesis, we depended on OD and its cross-validated counterpart, SD.cv in order to optimize Lasso SEM. This was based on the theoretical conclusion that OD can be used as a criteria for generalizability and reproducibility. However, other indices can replace them. For example, AIC or ECVI, which are considered as fit indices related to cross-validation and generalizability, can be exploited for optimization. They have advantage in that they can be used easily in practical data analysis.

Moreover, other optimization methods may lead us to better results. One candidate is a method based on estimated prediction error, which is described in *ii)* above. Another one is more closely related to SEM; penalize F discrepancy values as mentioned in limitation section. In this way, the optimization task will be carried out by minimizing $F(S, \hat{\Sigma}) + \kappa \times$ penalty. The MM-algorithm approach, another proposed algorithm for estimating Lasso SEM, also deserves our attention. In the present thesis, we pre-tested both of LARS and MM with a small-scale simulation. We couldn't find any considerable differences, but we selected the former as it consumed slightly less time. It is worthwhile conducting a more careful scrutiny so as to study the long-term performance of MM approach.

We investigated on several fit indices, but some of them including AIC, BIC, and ECVI require additional studies in a different research design. For example, the simulation should include a step in which multiple nested models with different degree of misspecification are fitted to 'one' fixed sample. And the models will be ordered by the observed size of those indices. What should

be inspected is the degree of agreement between the orders obtained by those indices and the orders acquired based on OD or sMSE. If information criteria or ECVI is able to yield great agreement with OD or sMSE, they can be evaluated as practical generalizability/reproducibility criteria. These indices can be examined accurately in this way since the important point of the indices is not their values perse but the orders they produce.

In order to check on the point we mentioned for BLasso SEM, comparing two models can gives us a meaningful implication. The one is a BLasso SEM model which derives some posterior distributions with their central tendencies close to zero. The other is almost the same model but has some differences. This model exclude some parameters from the beginning, which are corresponding to close-zero centered posterior distributions in the former model. Also this model is not analyzed with regularization; we fit the ordinary Bayesian SEM to this model. Therefore, the latter is similar to the 'hoped-for' result of the former BLasso SEM model, but it does not contain posterior distributions with close-zero central tendencies. If two models yields similar outcomes in terms of model fit, prediction, and other indices used in model evaluation, they can be regarded as essentially the same. However, when the opposite is the case, BLasso should face criticism on its shrinkage property.

Finally, other regularization methods, such as Elastic Net and Adaptive Lasso can also be applied to SEM. Since the Elastic Net exerts its shrinkage ability using both of L_1 and L_2 penalties, it may be able to produce a better result with respect to reducing OD and sMSE, maintaining the complete shrinkage result of Lasso. Adaptive Lasso is also able to draw Lasso's strength, with improved consistency of estimators. Those regularization methods might be easily applied by using the regularized conditional expectations of log-likelihoods of SEM which was derived in the present thesis.

References

Akaike, H. (1973). Information Theory and an Extension of the Maximum Likelihood Principle. In: Petrov, B.N., Csaki, F. (Eds.), *Proceeding of the Second International Symposium on Information Theory. Akademiai Kiado, Budapest*, 267–281.

Akaike, H. (1974). A New Look at the Statistical Model Identification. *IEEE Transactions on Automatic Control, 19*(6), 716-723.

Akaike, H. (1987). Factor Analysis and AIC. *Psychometrika, 52*(3), 317-332.

Andrews, D. F., & Mallows, C. L. (1974). Scale Mixtures of Normal Distributions. *Journal of the Royal Statistical Society: Series B (Methodological), 36*(1), 99-102.

Arbuckle, J. L. (2013). *IBM SPSS AMOS 22 User's Guide*. Armonk: IBM Corporation.

Bishop, C. (2006). *Pattern Recognition and Machine Learning.* New York: Springer. ISBN: 978-0-387-31073-2.

Bollen, K. A. (1989). *Structural Equations with Latent Variables.* New York: Wiley. ISBN: 978-0-471-01171-2.

Browne, M. W. (1973). Generalized Least Squares Estimators in the Analysis of Covariance Structures. *Research Bulletin, 1*, i-36.

Browne, M. W. (1984). Asymptotic Distribution Free Methods in Analysis of Covariance Structures. *British Journal of Mathematical and Statistical Psy-*

chology, 37(1), 62-83.

Browne, M. W., & Cudeck, R. (1989). Single Sample Cross-Validation Indices for Covariance Structures. *Multivariate Behavioral Research, 18*, 146-167.

Browne, M. W., & Cudeck, R. (1992). Alternative Ways of Assessing Model Fit. *Sociological Methods and Research, 24*(4), 445-455.

Browne, M. W. (2000). Cross-Validation Methods. *Journal of Mathematical Psychology, 44*, 108-132.

Browne, M. W., MacCallum, R. C., Kim, C., Andersen, B. L., & Glaser, R. (2002). When Fit Indices and Residuals Are Incompatible. *Psychological Methods, 7*(4), 403-421.

Browne, M. W., & MacCallum, R. C. (2004). Course Note on Covariance Structure Modeling. Unpublished Manuscript.

Cheung, G. W., & Rensvold, R. B. (2001). The Effects of Model Parsimony and Sampling Error on the Fit of Structural Equation Models. *Organizational Research Methods, 4*(3). 236-264.

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating Goodness-of-Fit Indexes for Testing Measurement Invariance. *Structural Equation Modeling: A Multidisciplinary Journal, 9*(2), 233-255.

Choi, J. (2010). Penalized Maximum Likelihood Factor Analysis. Ph. D. Dissertation, School of Statistics, Univ. of Minnesota.

Choi, J., Zou, H., & Oehlert, G. (2011). A Penalized Maximum Likelihood Approach to Sparse Factor Analysis. *Statistics and Its Interface, 3*(4), 429-436.

Chun, S. Y. & Shapiro, A. (2010). Construction of Covariance Matrices with a Specified Discrepancy Function Minimizer, with Application to Factor Analysis. *SIAM Journal on Matrix Analysis and Applications, 31*, 1570–1583.

Coffman, D. L. (2008). Model Error in Covariance Structure Models: Some Implications for Power and Type I Error. *Methodology, 4*(4). 159-167.

Cohen, J. (1962). The Statistical Power of Abnormal-Social Psychological Research: A Review. *Journal of Abnormal and Social Psychology, 65*(3), 145-153.

Cohen, J. (1992). Statistical Power Analysis. *Current Directions in Psychological Science, 1*(3), 98-101.

Cohen, J. (1994). The Earth Is Round($p < .05$). *American Psychologist Association, 49*(12), 997-1003.

Cudeck, R., & Henly, S. J. (1991). Model Selection in Covariance Structures Analysis and the Problem of Sample Size: A Clarification. *Psychological Bulletin, 109*(3), 512-519.

Cudeck, R., & Browne, M. W. (1983). Cross-Validation of Covariance Structures. *Multivariate Behavioral Research, 18*, 146-167.

Cudeck, R., & Browne, M. W. (1992). Constructing A Covariance Matrix That Yields A Specified Minimizer and A Specified Minimum Discrepancy Function Value. *Psychometrika, 57*(3), 357-369.

de Leeuw, J. (1992). Introduction to Akaike(1973) Information Theory and an Extension of the Maximum Likelihood Principle. In: Kotz, S., Johnson, N. L.(eds.) *Breakthroughs in Statistics Volume 1. Foundations and Basic Theory.* New York: Springer. 599–609.

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society, Series B 39*(1), 1–38.

Efron, B., Hastie, T., Johnstone, I., & Tibshirani, R. (2004). Least Angle Regression. *Annals of Statistics., 32*(2), 407–499.

Errington, T. M., Iorns, E., Gunn, W., Tan, F. E., Lomax, J., & Nosek, B. A. (2014). An Open Investigation of the Reproducibility of Cancer Biology Research. *eLife, 3*, e04333.

Francis, G. (2012). Publication Bias and the Failure of Replication in Experi-

mental Psychology. *Psychonomic Bulletin and Review, 19*, 975-991.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2014). *Bayesian Data Analysis*. Boca Raton: CRC Press.

Gill, J. (2007). *Bayesian Methods: A Social and Behavioral Sciences Approach.* New York : Chapman Hall/CRC.

Guo, R., Zhu, H., Chow, S., & Ibrahim, J. G. (2012). Bayesian Lasso for Semiparametric Structural Equation Models. *Biometrics, 68*, 567-577.

Hans, C. (2009). Bayesian Lasso Regression. *Biometrika, 96*(4), 835-845.

Hartshorne, J. K., & Schachner, A. (2012). Tracking Replicability as a Method of Post-publication Open Evaluation. *Frontiers in Computational Neuroscience, 6*(8), 1-14.

Hastie, T., Tibshirani, R., & Friedman, J. (2008). *The Elements of Statistical Learning; Data Mining, Inference, and Prediction.* New York: Springer. ISBN: 978-0387848570.

Hoerl, A. E., & Kennard, R. W. (1970a). Ridge Regression: Biased Estimation for Nonorthogonal Problems. *Technometrics, 12*(1), 55-67.

Hoerl, A. E., & Kennard, R. W. (1970b). Ridge Regression: Applications to Nonorthogonal Problems. *Technometrics, 12*(1), 55-67.

Hoff, P. D. (2009). *A First Course in Bayesian Statistical Methods.* New York: Springer. ISBN: 978-0-387-92299-7.

Hu, L., Bentler, P. M. (1998). Fit Indices in Covariance Structure Modeling: Sensitivity to Underparameterized Model Misspecification. *Psychological Methods, 3*(4), 424-453.

Hunter, D. R., & Lange, K. (2004). A Tutorial on MM Algorithms. *American Statistician, 58*(1), 30-37.

Hunter, D. R., & Li, R. (2005). Variable Selection Using MM Algorithms. *The Annals of Statistics, 33*(4), 1617-164237.

Ioannidis, J. P. A. (2005). Why Most Published Research Findings Are False. *PLoS Medicine, 2*, e124.

Jőreskog, K. G., & Sőrbom, D. (1993). *LISREL 8: Structural Equation Modeling with the SIMPLIS command language*. Lincolnwood, IL: Scientific Software International, Inc. ISBN: 0-89498-033-5

Jőreskog, K. G., & Sőrbom, D. (1996). *LISREL 8: User's Reference Guide*. Lincolnwood, IL: Scientific Software International, Inc. ISBN: 978-0894980404.

Kim, W. (2012). *Mathematical Statistics*. Seoul: Minyoungsa. ISBN: 978-89-8134-135-0. (In Korean)

Klein, J. G., Brown, G. T., & Lysyk, M. (2000). Replication Research: A Purposeful Occupation Worth Repeating. *The Canadian Journal of Occupational Therapy, 67*(3), 155-161.

Klein, R. A., Ratliff, K. A., Vianello, M., Adams, R. B., Bahnik, S., Bernstein, M. J., . . . Nosek, B. A. (2014). Investigating Variation in Replicability: A "Many Labs" Replication Project. *Social Psychology, 45*(3), 142–152. doi: 10.1027/1864-9335/a000178

Koole, S. L., & Lakens, D. (2012). Rewarding Replications: A Sure and Simple Way to Improve Psychological Science. *Perspectives on Psychological Science, 7*(6), 608-614.

Kyung, M. (2014). Comparison of Laplace and Double Pareto Penalty: Lasso 와 Elastic Net. *The Korean Journal of Applied Statistics, 27*(6), 975-989. (In Korean)

Lange, K. (2013) *Optimization.* New York: Springer. ISBN: 978-1-4614-5831-1.

LeBel, E. P., & Paunonen, S. V. (2011). Sexy But Often Unreliable: The Impact of Unreliability on the Replicability of Experimental Findings With Implicit Measures. *Personality and Social Psychology Bulletin, 37*(4), 570–583.

Lee, S.-Y. (2007). *Structural Equation Modeling : A Bayesian Approach*. New York:

Wiley. ISBN: 978-0-470-02423-2.

MacCallum, R. C., Roznowski, M., Mar, C. M., & Reith, J. V. (1994). Alternative Strategies for Cross-Validation of Covariance Structure Models, *Multivariate Behavioral Research, 29*(1), 1-32.

MacCallum, R. C., Lee, T., & Browne, M. W. (2010). The Issue of Isopower in Power Analysis for Tests of Structural Equation Models. *Structural Equation Modeling: A Multidisciplinary Journal, 17*, 23-41.

Magnus, J. R., Neudecker, H. (1999). *Matrix Differential Calculus with Applications in Statistics and Econometrics*. New York: Wiley. ISBN: 978-0-471-98633-1.

Makel, M. C., Plucker, J. A., & Hegarty, B. (2012). Replications in Psychology Research: How Often Do They Really Occur? *Perspectives on Psychological Science, 7*(6), 537-542.

McDonald, R. P. (1989). An Index of Goodness-of-Fit Based on Noncentrality. *Journal of Classification, 6*, 97-103.

Montgomery, D. C., Peck, E. A., Vining, G. G. (2012). *Introduction to Linear Regression Analysis*. New York: Wiley. ISBN: 978-0-470-54281-1.

Myung, J. I., & Pitt, M. A. (1997). Applying Occam's razor in modeling cognition: A Bayesian approach. *Psychonomic Bulletin & Review, 4*(1), 79-95.

Myung, J. I. (2000). The Importance of Complexity in Model Selection *Journal of Mathematical Psychology, 44*, 190-204.

Nel, D. G. (1980). On Matrix Differentiation In Statistics. *South African Statistics Journal, 14*, 137-193.

Neuliep, J. W., & Crandall, R. (1990). Editorial Bias Against Replication Research. *Journal of Social Behavior and Personality, 5*(4), 85-90.

Nosek, B. A., & Bar-Anan, Y. (2012). Scientific Utopia: I. Opening Scientific Communication. *Psychological Inquiry, 23*, 217-243.

Nosek, B. A., Spies, J. R., & Motyl, M. (2012). Scientific Utopia: II. Restructuring Incentives and Practices to Promote Truth Over Publishability. *Perspectives on Psychological Science, 7*(6), 615–631.

Olsson, U. H., Foss, T., & Breivik, E. (2014). Two Equivalent Discrepancy Functions for Maximum Likelihood Estimation: Do Their Test Statistics Follow a Non-central Chi-square Distribution under Model Misspecification?. *Sociological Methods & Research, 32*(4), 453-500.

Open Science Collaboration. (2012). An Open, Large-Scale, Collaborative Effort to Estimate the Reproducibility of Psychological Science. *Perspectives on Psychological Science, 7*(6), 657-660.

Open Science Collaboration. (2014). The reproducibility project: A Model of Large-Scale Collaboration for Empirical Research on Reproducibility. In: Stodden, V., Leisch, F., Peng, R. (eds.) *Implementing Reproducible Computational Research (A Volume in the R Series)*. New York: Taylor Francis. 299–323.

Open Science Collaboration. (2015). Estimating the Reproducibility of Psychological Science. *Science 349*, aac4716. doi: 10.1126/science.aac4716.

Ostrowsky, A. M. (1959). A Quantitative Formulation of Sylvester's Law of Inertia. *Proceedings of the National Academy of Sciences, USA, 45*, 740-744.

Paik, M. C. (2014). Ch13 : MM algorithm [PowerPoint slides]. A Lecture Note on 'Applied Statistics', Department of Statistics, Seoul National University.

Park, T., & Casella, G. (2008). The Bayesian Lasso. *Journal of the American Statistical Association, 103*(482), 681-686.

Preacher, K. J., Zhang, G., Kim, C., & Mels, G. (2013). Choosing the Optimal Number of Factors in Exploratory Factor Analysis: A Model Selection Perspective. *Multivariate Behavioral Research, 48*, 28-56.

Rossi, J. S. (1990). Statistical Power of Psychological Research: What Have We Gained in 20 Years? *Journal of Consulting and Clinical Psychology, 58*(5), 646-656.

Schwarz, G. (1978). Estimating the Dimension of a Model. *The Annals of Statistics, 6*(2), 461-464.

Schmidt, F. L., & Hunter, J. E. (1996). Measurement Error in Psychological Research: Lessons From 26 Research Scenarios.] *Psychological Methods, 1*(2), 199-223.

Schmidt, S. (2009). Shall We Really Do It Again? The Powerful Concept of Replication Is Neglected in the Social Sciences. *Review of General Psychology, 13*(2), 90-100.

Sedlmeier, P., & Gigerenzer, G. (1989). Do Studies of Statistical Power Have an Effect on the Power of Studies? *Psychological Bulletin, 105*(2), 309-316

Simmons J. P., Nelson, L. D., & Simonsohn, U. (2011). False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant. *Psychological Science, 22*(11), 1359-1366.

Song, X.-Y., & Lee, S.-Y. (2012a). A tutorial on the Bayesian approach for analyzing structural equation models. *Journal of Mathematical Psychology, 56*, 135-148.

Song, X.-Y., & Lee, S.-Y. (2012b). *Basic and Advanced Bayesian Structural Equation Modeling: With Applications in the Medical and Behavioral Sciences.* New York: Wiley. ISBN: 978-0-470-66952-5.

Steiger, J. H., & Lind, J. C. (1980). Statistically Based Tests for the Number of Common Factors. Paper presented at the annual meeting of the Psychometric Society, Iowa City.

Steiger, J. H., Shapiro, A., & Browne, M. W. (1985). On the Multivariate Asymptotic Distribution of Sequential Chi-square Statistics. *Psychometrika, 50*(3),

253-264.

Tibshirani, R. (1996). Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society: Series B (Methodological), 58*(1), 267-288.

Tomarken, A. J., & Waller, N. G. (2005). Structural Equation Modeling: Strengths, Limitations, and Misconceptions. *Annual Review of Clinical Psychology, 1*, 31-65.

Tucker, L. R., Koopman, R. F., & Linn, R. L. (1969). Evaluation of Factor Analytic Research Procedures by means of Simulated Correlation Matrices. *Psychometrika, 34*, 421-459.

Wang, Z. (2014). Bayesian Non-Linear Methods for Survival Analysis and Structural Equation Models. Ph. D. Dissertation, School of Statistics, Univ. of Missouri.

Wasserman, L. (2004). *All of Statistics: A Concise Course in Statistical Inference.* New York: Springer. ISBN: 978-0-387-40272-7.

Wasserman, L. (2006). *All of Nonparametric Statistics.* New York: Springer. ISBN: 978-0-387-30623-0.

Yuan, K.-H., Wu, R., & Bentler, P. M. (2010). Ridge Structural Equation Modelling with Correlation Matrices for Ordinal and Continuous data. *British Journal of Mathematical and Statistical Psychology, 64*, 107-133.

Zou, H., & Hastie, T. (2005). Regularization and Variable Selection via the Elastic Net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 67*(2), 301-320.

Appendix

Appendix A : Result Tables

A1. Result Tables for Lasso

This section provides result tables not presented in the previous chapters. This covers the Lasso outcomes including DE, MSE, and Variance. Table A1.1∼ A1.6 analyze those indices observed in Lasso FA simulations. The other tables show Lasso SEM results. Some notable results can be summarized as follows.

- Lasso SEM shows outstanding performance on reducing DE, whether the model error is employed or not. Thus, the proposed method can produce stable results over several independent samples. But its larger DA devaluates this result.
- MSE follows a similar tendency to sMSE in general. Rigorously speaking, it is slightly less supportive for L_1-regularized SEM, especially in SEM with model error involved case.
- As DE, Lasso seems to be able to diminish variances of parameters. However, unlike DE, this performance is severely influenced by the degree of misspecification. This is equal to what we discuss about differences of OD and sMSE in Chapter 6.
- The trend of variance reduction is also similar to that of sMSE; Lasso SEM produces great results in Φ_1 and Φ_2 conditions.

DE : Discrepancy due to Estimation (OLS)

Sample Size	Φ	Model 1					Model 2				
		ML	Lasso		Lasso (OD)		ML	Lasso		Lasso (OD)	
50	Φ_1	0.9443	0.6416	(0.82)	0.6173	(0.91)	1.0040	0.7229	(0.85)	0.6925	(0.92)
	Φ_2	0.7676	0.5928	(0.84)	0.5761	(0.90)	0.7377	0.5585	(0.84)	0.5386	(0.95)
	Φ_3	0.7939	0.5758	(0.77)	0.5554	(0.80)	0.7790	0.5374	(0.74)	0.5144	(0.81)
	Φ_4	0.6764	0.5018	(0.78)	0.4871	(0.85)	0.6597	0.5422	(0.66)	0.5264	(0.72)
	Φ_5	0.6738	0.6212	(0.38)	0.6045	(0.41)	0.5618	0.6204	(0.34)	0.6075	(0.36)
	Φ_6	0.5401	0.5865	(0.29)	0.5751	(0.33)	0.5226	0.5245	(0.38)	0.5164	(0.40)
100	Φ_1	0.3873	0.2779	(0.84)	0.2678	(0.96)	0.4576	0.3175	(0.76)	0.2997	(0.92)
	Φ_2	0.3550	0.2801	(0.76)	0.2700	(0.89)	0.3764	0.2914	(0.88)	0.2850	(0.97)
	Φ_3	0.3950	0.2553	(0.80)	0.2424	(0.83)	0.3374	0.2642	(0.65)	0.2469	(0.80)
	Φ_4	0.3390	0.2436	(0.73)	0.2341	(0.80)	0.2972	0.2235	(0.72)	0.2144	(0.81)
	Φ_5	0.2872	0.3000	(0.40)	0.2920	(0.40)	0.3091	0.3392	(0.30)	0.3308	(0.33)
	Φ_6	0.2748	0.3085	(0.31)	0.3017	(0.33)	0.2889	0.3247	(0.35)	0.3176	(0.38)
200	Φ_1	0.2329	0.1443	(0.85)	0.1365	(0.95)	0.2222	0.1291	(0.86)	0.1198	(0.94)
	Φ_2	0.2089	0.1644	(0.81)	0.1600	(0.93)	0.2020	0.1557	(0.88)	0.1518	(0.97)
	Φ_3	0.1653	0.1230	(0.57)	0.1137	(0.74)	0.1860	0.1255	(0.69)	0.1203	(0.72)
	Φ_4	0.1315	0.1067	(0.60)	0.1010	(0.70)	0.1815	0.1332	(0.72)	0.1281	(0.84)
	Φ_5	0.1604	0.1602	(0.35)	0.1560	(0.37)	0.1584	0.1731	(0.36)	0.1681	(0.40)
	Φ_6	0.1326	0.1536	(0.28)	0.1512	(0.29)	0.1373	0.1539	(0.37)	0.1524	(0.37)
1000	Φ_1	0.0475	0.0241	(0.79)	0.0222	(0.93)	0.0440	0.0247	(0.83)	0.0228	(0.95)
	Φ_2	0.0316	0.0218	(0.81)	0.0206	(0.93)	0.0354	0.0240	(0.82)	0.0228	(0.95)
	Φ_3	0.0360	0.0243	(0.63)	0.0223	(0.74)	0.0307	0.0215	(0.66)	0.0197	(0.75)
	Φ_4	0.0310	0.0228	(0.76)	0.0218	(0.79)	0.0305	0.0228	(0.71)	0.0215	(0.81)
	Φ_5	0.0294	0.0299	(0.33)	0.0290	(0.35)	0.0327	0.0367	(0.31)	0.0360	(0.33)
	Φ_6	0.0254	0.0298	(0.28)	0.0293	(0.29)	0.0273	0.0306	(0.39)	0.0301	(0.39)

Sample Size	Φ	Model 3					Model 4				
		ML	Lasso		Lasso (OD)		ML	Lasso		Lasso (OD)	
50	Φ_1	1.0567	0.6768	(0.92)	0.6611	(0.95)	1.0787	0.6518	(0.88)	0.6286	(0.94)
	Φ_2	0.8863	0.6598	(0.91)	0.6462	(0.96)	0.7471	0.6012	(0.91)	0.5861	(0.98)
	Φ_3	0.8039	0.6277	(0.58)	0.6065	(0.62)	0.9087	0.5944	(0.74)	0.5734	(0.83)
	Φ_4	0.6776	0.5441	(0.67)	0.5272	(0.74)	0.6764	0.5365	(0.84)	0.5246	(0.86)
	Φ_5	0.6340	0.6676	(0.29)	0.6547	(0.29)	0.6996	0.6942	(0.40)	0.6795	(0.41)
	Φ_6	0.6581	0.6709	(0.41)	0.6621	(0.44)	0.6566	0.6288	(0.45)	0.6172	(0.48)
100	Φ_1	0.4565	0.3179	(0.82)	0.3061	(0.92)	0.5584	0.3445	(0.91)	0.3265	(0.98)
	Φ_2	0.4251	0.2884	(0.92)	0.2831	(0.99)	0.4210	0.2929	(0.91)	0.2847	(0.97)
	Φ_3	0.3734	0.2898	(0.72)	0.2760	(0.73)	0.4489	0.2937	(0.79)	0.2791	(0.87)
	Φ_4	0.3696	0.2730	(0.78)	0.2639	(0.80)	0.3504	0.2503	(0.84)	0.2401	(0.90)
	Φ_5	0.3164	0.4150	(0.22)	0.4072	(0.26)	0.3438	0.3339	(0.53)	0.3262	(0.57)
	Φ_6	0.2973	0.3186	(0.43)	0.3104	(0.44)	0.3018	0.3026	(0.46)	0.2962	(0.47)
200	Φ_1	0.2379	0.1497	(0.83)	0.1439	(0.94)	0.3010	0.1820	(0.94)	0.1751	(0.98)
	Φ_2	0.2048	0.1449	(0.93)	0.1384	(0.99)	0.2213	0.1496	(0.89)	0.1450	(0.96)
	Φ_3	0.2109	0.1503	(0.63)	0.1413	(0.74)	0.2283	0.1391	(0.87)	0.1324	(0.91)
	Φ_4	0.1716	0.1242	(0.85)	0.1189	(0.89)	0.1933	0.1297	(0.89)	0.1255	(0.92)
	Φ_5	0.1722	0.2776	(0.23)	0.2735	(0.25)	0.1599	0.1573	(0.42)	0.1538	(0.45)
	Φ_6	0.1572	0.1710	(0.41)	0.1691	(0.46)	0.1677	0.1652	(0.49)	0.1616	(0.53)
1000	Φ_1	0.0449	0.0251	(0.85)	0.0230	(0.93)	0.0417	0.0227	(0.88)	0.0199	(0.99)
	Φ_2	0.0398	0.0265	(0.88)	0.0249	(0.95)	0.0460	0.0281	(0.89)	0.0271	(0.95)
	Φ_3	0.0363	0.0266	(0.59)	0.0246	(0.64)	0.0433	0.0223	(0.89)	0.0206	(0.94)
	Φ_4	0.0356	0.0236	(0.89)	0.0225	(0.96)	0.0354	0.0236	(0.86)	0.0224	(0.91)
	Φ_5	0.0334	0.0679	(0.14)	0.0670	(0.17)	0.0303	0.0312	(0.39)	0.0303	(0.41)
	Φ_6	0.0295	0.0296	(0.51)	0.0290	(0.51)	0.0306	0.0303	(0.51)	0.0298	(0.52)

Table A1.1: DE Table - Lasso FA, No Model Error

DE : Discrepancy due to Estimation (OLS)

Sample Size	Φ	Model 1					Model 2				
		ML	Lasso		Lasso (OD)		ML	Lasso		Lasso (OD)	
50	Φ_1	1.0613	0.7054	(0.60)	0.6854	(0.63)	0.9484	0.6379	(0.61)	0.6079	(0.71)
	Φ_2	0.8672	0.5768	(0.76)	0.5616	(0.82)	0.8511	0.5981	(0.74)	0.5870	(0.78)
	Φ_3	0.7355	0.5625	(0.72)	0.5444	(0.78)	0.7578	0.5768	(0.69)	0.5534	(0.74)
	Φ_4	0.6408	0.5415	(0.71)	0.5218	(0.80)	0.6974	0.5309	(0.81)	0.5164	(0.84)
	Φ_5	0.6522	0.6046	(0.45)	0.5958	(0.47)	0.6182	0.7848	(0.34)	0.7649	(0.32)
	Φ_6	0.5730	0.5672	(0.45)	0.5558	(0.46)	0.6049	0.6258	(0.44)	0.6141	(0.44)
100	Φ_1	0.4537	0.3449	(0.75)	0.3284	(0.87)	0.4592	0.3115	(0.83)	0.3009	(0.87)
	Φ_2	0.3557	0.2861	(0.83)	0.2760	(0.91)	0.3898	0.2931	(0.80)	0.2851	(0.89)
	Φ_3	0.3699	0.2555	(0.72)	0.2427	(0.78)	0.3879	0.2701	(0.68)	0.2594	(0.74)
	Φ_4	0.3704	0.2862	(0.71)	0.2793	(0.77)	0.3179	0.2424	(0.70)	0.2345	(0.79)
	Φ_5	0.3076	0.2980	(0.45)	0.2910	(0.49)	0.3187	0.4979	(0.21)	0.5082	(0.16)
	Φ_6	0.2652	0.2818	(0.39)	0.2754	(0.41)	0.2767	0.3165	(0.27)	0.3091	(0.29)
200	Φ_1	0.2171	0.1449	(0.84)	0.1360	(0.94)	0.2388	0.1581	(0.83)	0.1495	(0.92)
	Φ_2	0.2174	0.1453	(0.82)	0.1409	(0.93)	0.1888	0.1419	(0.77)	0.1363	(0.91)
	Φ_3	0.1889	0.1217	(0.69)	0.1154	(0.81)	0.1959	0.1318	(0.61)	0.1261	(0.68)
	Φ_4	0.1729	0.1229	(0.74)	0.1181	(0.80)	0.1674	0.1335	(0.63)	0.1266	(0.76)
	Φ_5	0.1434	0.1591	(0.31)	0.1560	(0.31)	0.1468	0.3591	(0.11)	0.3612	(0.10)
	Φ_6	0.1297	0.1461	(0.40)	0.1439	(0.41)	0.1620	0.1800	(0.38)	0.1770	(0.41)
1000	Φ_1	0.0446	0.0250	(0.82)	0.0236	(0.90)	0.0397	0.0235	(0.76)	0.0214	(0.91)
	Φ_2	0.0397	0.0267	(0.83)	0.0257	(0.90)	0.0400	0.0279	(0.79)	0.0266	(0.89)
	Φ_3	0.0381	0.0241	(0.68)	0.0231	(0.64)	0.0377	0.0243	(0.73)	0.0234	(0.72)
	Φ_4	0.0290	0.0216	(0.72)	0.0204	(0.76)	0.0302	0.0237	(0.63)	0.0227	(0.70)
	Φ_5	0.0313	0.0466	(0.17)	0.0453	(0.14)	0.0295	0.2983	(0.00)	0.3033	(0.00)
	Φ_6	0.0260	0.0289	(0.36)	0.0283	(0.36)	0.0295	0.0345	(0.32)	0.0341	(0.34)

Sample Size	Φ	Model 3					Model 4				
		ML	Lasso		Lasso (OD)		ML	Lasso		Lasso (OD)	
50	Φ_1	0.9392	0.6918	(0.76)	0.6689	(0.81)	1.1087	0.7399	(0.68)	0.7119	(0.72)
	Φ_2	0.9552	0.7368	(0.85)	0.7178	(0.89)	0.8647	0.6793	(0.79)	0.6592	(0.87)
	Φ_3	0.7192	0.6013	(0.55)	0.5751	(0.63)	0.8667	0.6550	(0.84)	0.6381	(0.88)
	Φ_4	0.7970	0.6403	(0.72)	0.6214	(0.79)	0.7082	0.5696	(0.79)	0.5518	(0.86)
	Φ_5	0.6507	0.8049	(0.27)	0.7956	(0.29)	0.7147	0.7082	(0.48)	0.6972	(0.50)
	Φ_6	0.6094	0.6963	(0.29)	0.6839	(0.30)	0.6517	0.6294	(0.48)	0.6229	(0.50)
100	Φ_1	0.5571	0.3480	(0.80)	0.3339	(0.92)	0.5628	0.3857	(0.88)	0.3724	(0.94)
	Φ_2	0.3582	0.2987	(0.81)	0.2883	(0.94)	0.4256	0.3356	(0.84)	0.3268	(0.92)
	Φ_3	0.3655	0.2988	(0.58)	0.2889	(0.65)	0.4600	0.3065	(0.82)	0.2945	(0.89)
	Φ_4	0.3817	0.2962	(0.70)	0.2880	(0.76)	0.3768	0.2866	(0.81)	0.2757	(0.87)
	Φ_5	0.3137	0.5604	(0.20)	0.5431	(0.22)	0.3458	0.3279	(0.49)	0.3188	(0.56)
	Φ_6	0.3324	0.3865	(0.28)	0.3759	(0.33)	0.3326	0.3456	(0.40)	0.3408	(0.41)
200	Φ_1	0.2435	0.1633	(0.81)	0.1569	(0.92)	0.2945	0.1922	(0.91)	0.1858	(0.98)
	Φ_2	0.2216	0.1738	(0.85)	0.1706	(0.92)	0.2367	0.1653	(0.88)	0.1616	(0.98)
	Φ_3	0.2027	0.1540	(0.60)	0.1480	(0.68)	0.2082	0.1338	(0.88)	0.1268	(0.96)
	Φ_4	0.1835	0.1445	(0.65)	0.1389	(0.68)	0.1780	0.1341	(0.78)	0.1279	(0.87)
	Φ_5	0.1763	0.4237	(0.26)	0.4212	(0.23)	0.1784	0.1855	(0.40)	0.1810	(0.42)
	Φ_6	0.1489	0.1999	(0.28)	0.1972	(0.30)	0.1539	0.1524	(0.48)	0.1497	(0.52)
1000	Φ_1	0.0462	0.0247	(0.82)	0.0230	(0.92)	0.0463	0.0259	(0.87)	0.0238	(0.98)
	Φ_2	0.0407	0.0280	(0.80)	0.0267	(0.93)	0.0388	0.0263	(0.95)	0.0254	(0.98)
	Φ_3	0.0368	0.0269	(0.60)	0.0254	(0.67)	0.0436	0.0248	(0.77)	0.0234	(0.81)
	Φ_4	0.0393	0.0304	(0.66)	0.0296	(0.69)	0.0343	0.0225	(0.84)	0.0214	(0.88)
	Φ_5	0.0351	0.2322	(0.05)	0.2346	(0.04)	0.0345	0.0425	(0.19)	0.0430	(0.16)
	Φ_6	0.0290	0.0480	(0.16)	0.0466	(0.17)	0.0290	0.0301	(0.45)	0.0295	(0.47)

Table A1.2: DE Table - Lasso FA, Model Error Involved

MSE : Mean Squared Error

Sample Size	Φ	Model 1					Model 2				
		ML	Lasso		Lasso (OD)		ML	Lasso		Lasso (OD)	
50	Φ_1	0.2259	0.2120	(0.64)	0.2047	(0.71)	0.2797	0.2476	(0.63)	0.2425	(0.67)
	Φ_2	0.2161	0.2051	(0.67)	0.2018	(0.70)	0.2136	0.1991	(0.65)	0.1954	(0.69)
	Φ_3	0.4939	0.5833	(0.26)	0.5579	(0.28)	0.6492	0.6606	(0.41)	0.6587	(0.42)
	Φ_4	0.5157	0.6873	(0.19)	0.6737	(0.20)	0.5825	0.7121	(0.26)	0.6983	(0.23)
	Φ_5	1.1269	2.3784	(0.05)	2.2398	(0.04)	6.2839	11.2763	(0.23)	11.6140	(0.23)
	Φ_6	2.1232	4.7034	(0.06)	4.4213	(0.06)	2.3944	3.2630	(0.04)	3.2199	(0.04)
100	Φ_1	0.1159	0.1076	(0.54)	0.1061	(0.61)	0.1314	0.1171	(0.67)	0.1128	(0.70)
	Φ_2	0.1008	0.0961	(0.59)	0.0941	(0.66)	0.1150	0.1066	(0.67)	0.1056	(0.70)
	Φ_3	0.2155	0.2572	(0.23)	0.2527	(0.22)	0.3020	0.3534	(0.31)	0.3438	(0.32)
	Φ_4	0.2271	0.3045	(0.23)	0.2980	(0.22)	0.2342	0.3023	(0.19)	0.2928	(0.19)
	Φ_5	0.5207	1.0256	(0.06)	0.9953	(0.04)	2.1160	1.1750	(0.24)	1.1537	(0.20)
	Φ_6	0.5663	1.4676	(0.01)	1.4438	(0.01)	0.6424	1.4139	(0.05)	1.3567	(0.05)
200	Φ_1	0.0544	0.0482	(0.56)	0.0476	(0.58)	0.0587	0.0502	(0.61)	0.0496	(0.59)
	Φ_2	0.0562	0.0546	(0.58)	0.0538	(0.63)	0.0578	0.0551	(0.58)	0.0540	(0.59)
	Φ_3	0.1035	0.1372	(0.13)	0.1342	(0.13)	0.1304	0.1718	(0.20)	0.1722	(0.20)
	Φ_4	0.0983	0.1274	(0.16)	0.1271	(0.16)	0.1215	0.1498	(0.22)	0.1475	(0.23)
	Φ_5	0.2461	0.5037	(0.04)	0.4939	(0.04)	0.3721	0.6937	(0.19)	0.6945	(0.18)
	Φ_6	0.2590	0.6219	(0.01)	0.6213	(0.01)	0.2793	0.6494	(0.02)	0.6407	(0.02)
1000	Φ_1	0.0114	0.0110	(0.40)	0.0108	(0.41)	0.0126	0.0132	(0.37)	0.0129	(0.37)
	Φ_2	0.0093	0.0102	(0.35)	0.0099	(0.41)	0.0104	0.0107	(0.39)	0.0105	(0.43)
	Φ_3	0.0210	0.0343	(0.04)	0.0342	(0.04)	0.0237	0.0481	(0.06)	0.0479	(0.06)
	Φ_4	0.0207	0.0342	(0.06)	0.0335	(0.07)	0.0210	0.0334	(0.07)	0.0331	(0.08)
	Φ_5	0.0416	0.1131	(0.00)	0.1131	(0.00)	0.0714	0.1835	(0.09)	0.1826	(0.09)
	Φ_6	0.0467	0.1492	(0.00)	0.1492	(0.00)	0.0496	0.1444	(0.00)	0.1442	(0.00)

Sample Size	Φ	Model 3					Model 4				
		ML	Lasso		Lasso (OD)		ML	Lasso		Lasso (OD)	
50	Φ_1	0.4715	0.3446	(0.80)	0.3430	(0.82)	0.3145	0.2710	(0.78)	0.2692	(0.83)
	Φ_2	0.3013	0.2620	(0.82)	0.2621	(0.81)	0.2576	0.2421	(0.76)	0.2384	(0.80)
	Φ_3	1.5492	1.1886	(0.56)	1.1841	(0.53)	0.5417	0.6195	(0.27)	0.6094	(0.28)
	Φ_4	0.7387	0.8011	(0.36)	0.7772	(0.34)	0.6411	0.7740	(0.17)	0.7424	(0.19)
	Φ_5	17.1535	16.5423	(0.43)	16.6399	(0.43)	1.5822	3.4124	(0.10)	2.7283	(0.07)
	Φ_6	5.8527	8.7349	(0.14)	9.0465	(0.13)	1.7714	3.8126	(0.05)	5.3157	(0.06)
100	Φ_1	0.2286	0.1747	(0.84)	0.1685	(0.85)	0.1479	0.1248	(0.80)	0.1243	(0.80)
	Φ_2	0.1314	0.1107	(0.85)	0.1094	(0.86)	0.1288	0.1181	(0.73)	0.1156	(0.79)
	Φ_3	0.6820	0.5441	(0.49)	0.5400	(0.49)	0.2551	0.2881	(0.35)	0.2819	(0.38)
	Φ_4	0.3287	0.3398	(0.40)	0.3417	(0.38)	0.2514	0.2949	(0.33)	0.2917	(0.34)
	Φ_5	45.5912	3.3495	(0.34)	3.2642	(0.35)	0.5202	1.0061	(0.05)	1.0090	(0.05)
	Φ_6	0.8856	1.4610	(0.20)	1.4053	(0.19)	0.6415	1.5311	(0.03)	1.5137	(0.03)
200	Φ_1	0.0969	0.0714	(0.75)	0.0700	(0.81)	0.0810	0.0675	(0.76)	0.0659	(0.79)
	Φ_2	0.0635	0.0558	(0.76)	0.0542	(0.82)	0.0741	0.0663	(0.72)	0.0658	(0.76)
	Φ_3	0.2875	0.3310	(0.33)	0.3241	(0.33)	0.1262	0.1443	(0.27)	0.1436	(0.28)
	Φ_4	0.1383	0.1502	(0.46)	0.1472	(0.51)	0.1346	0.1572	(0.31)	0.1528	(0.29)
	Φ_5	1.4564	2.1015	(0.32)	2.1083	(0.32)	0.2540	0.5372	(0.04)	0.5301	(0.04)
	Φ_6	0.3905	0.8205	(0.10)	0.8124	(0.09)	0.2561	0.6234	(0.04)	0.6162	(0.04)
1000	Φ_1	0.0192	0.0170	(0.64)	0.0168	(0.64)	0.0124	0.0122	(0.46)	0.0118	(0.46)
	Φ_2	0.0124	0.0110	(0.62)	0.0109	(0.63)	0.0139	0.0130	(0.56)	0.0128	(0.59)
	Φ_3	0.0538	0.0975	(0.17)	0.0979	(0.17)	0.0229	0.0315	(0.16)	0.0311	(0.15)
	Φ_4	0.0290	0.0356	(0.29)	0.0355	(0.29)	0.0240	0.0356	(0.16)	0.0356	(0.14)
	Φ_5	0.1598	0.5683	(0.13)	0.5749	(0.13)	0.0472	0.1305	(0.00)	0.1305	(0.00)
	Φ_6	0.0672	0.1510	(0.04)	0.1496	(0.04)	0.0499	0.1523	(0.01)	0.1503	(0.01)

Table A1.3: MSE Table - Lasso FA, No Model Error

MSE : Mean Squared Error

Sample Size	Φ	Model 1					Model 2				
		ML	Lasso		Lasso (OD)		ML	Lasso		Lasso (OD)	
50	Φ_1	0.2713	0.2446	(0.62)	0.2414	(0.68)	0.2602	0.2401	(0.56)	0.2322	(0.60)
	Φ_2	0.2326	0.2162	(0.67)	0.2148	(0.67)	0.2345	0.2240	(0.67)	0.2229	(0.68)
	Φ_3	0.4964	0.5715	(0.22)	0.5713	(0.26)	0.6853	0.6808	(0.30)	0.6630	(0.33)
	Φ_4	0.4839	0.6284	(0.19)	0.5987	(0.21)	0.6364	0.7158	(0.31)	0.7083	(0.33)
	Φ_5	1.1198	2.3581	(0.11)	2.3906	(0.11)	6.6659	3.2571	(0.30)	3.3436	(0.29)
	Φ_6	1.4877	2.8655	(0.06)	2.8370	(0.07)	5.0454	9.0030	(0.09)	8.2016	(0.10)
100	Φ_1	0.1070	0.0955	(0.74)	0.0922	(0.78)	0.1346	0.1276	(0.44)	0.1264	(0.52)
	Φ_2	0.1027	0.0977	(0.64)	0.0955	(0.69)	0.1126	0.1024	(0.75)	0.1013	(0.77)
	Φ_3	0.2248	0.2791	(0.21)	0.2669	(0.21)	0.3008	0.3616	(0.20)	0.3557	(0.20)
	Φ_4	0.2556	0.3103	(0.16)	0.3057	(0.17)	0.2508	0.3046	(0.26)	0.3005	(0.26)
	Φ_5	0.5835	1.0485	(0.08)	1.0434	(0.06)	2.8097	1.5380	(0.17)	1.4944	(0.17)
	Φ_6	0.5490	1.2791	(0.02)	1.1985	(0.03)	0.7177	1.7204	(0.05)	1.5925	(0.04)
200	Φ_1	0.0514	0.0462	(0.61)	0.0443	(0.64)	0.0737	0.0713	(0.46)	0.0702	(0.47)
	Φ_2	0.0558	0.0509	(0.56)	0.0499	(0.60)	0.0553	0.0522	(0.61)	0.0512	(0.66)
	Φ_3	0.1006	0.1287	(0.15)	0.1267	(0.14)	0.1666	0.2453	(0.12)	0.2437	(0.14)
	Φ_4	0.1172	0.1416	(0.25)	0.1397	(0.27)	0.1288	0.1687	(0.08)	0.1668	(0.10)
	Φ_5	0.2173	0.4767	(0.03)	0.4647	(0.04)	0.4051	1.1030	(0.07)	1.0919	(0.07)
	Φ_6	0.2893	0.6633	(0.02)	0.6578	(0.02)	0.3307	0.8827	(0.01)	0.8537	(0.01)
1000	Φ_1	0.0107	0.0100	(0.46)	0.0098	(0.52)	0.0192	0.0251	(0.09)	0.0250	(0.09)
	Φ_2	0.0101	0.0100	(0.51)	0.0099	(0.49)	0.0150	0.0140	(0.60)	0.0140	(0.60)
	Φ_3	0.0218	0.0413	(0.02)	0.0406	(0.03)	0.0446	0.1147	(0.00)	0.1139	(0.00)
	Φ_4	0.0194	0.0355	(0.04)	0.0352	(0.05)	0.0275	0.0455	(0.02)	0.0454	(0.02)
	Φ_5	0.0430	0.1815	(0.00)	0.1824	(0.00)	0.0887	0.4658	(0.00)	0.4544	(0.00)
	Φ_6	0.0461	0.1768	(0.00)	0.1754	(0.00)	0.0646	0.2253	(0.00)	0.2259	(0.00)

Sample Size	Φ	Model 3					Model 4				
		ML	Lasso		Lasso (OD)		ML	Lasso		Lasso (OD)	
50	Φ_1	0.4653	0.3687	(0.72)	0.3668	(0.72)	0.2888	0.2571	(0.78)	0.2542	(0.86)
	Φ_2	0.2972	0.2754	(0.77)	0.2714	(0.81)	0.2749	0.2498	(0.80)	0.2471	(0.80)
	Φ_3	1.2897	1.2965	(0.35)	1.2682	(0.34)	0.5328	0.5948	(0.27)	0.5983	(0.27)
	Φ_4	0.7447	0.8510	(0.28)	0.8364	(0.26)	0.6107	0.8435	(0.18)	0.7875	(0.19)
	Φ_5	173.6098	56.9230	(0.36)	56.6222	(0.36)	1.2074	2.4819	(0.13)	2.4375	(0.15)
	Φ_6	3.2013	7.7950	(0.14)	7.8200	(0.15)	3.0894	274.0369	(0.12)	274.2774	(0.13)
100	Φ_1	0.2268	0.1871	(0.72)	0.1904	(0.71)	0.1572	0.1371	(0.83)	0.1353	(0.88)
	Φ_2	0.1412	0.1251	(0.81)	0.1243	(0.82)	0.1488	0.1388	(0.76)	0.1372	(0.75)
	Φ_3	0.6364	0.8194	(0.19)	0.7963	(0.19)	0.2629	0.2989	(0.28)	0.2968	(0.28)
	Φ_4	0.3980	0.4597	(0.29)	0.4522	(0.28)	0.2618	0.3033	(0.27)	0.3016	(0.27)
	Φ_5	3.2446	3.1947	(0.18)	3.2581	(0.18)	0.5064	0.9421	(0.10)	0.9292	(0.09)
	Φ_6	1.0266	1.9109	(0.09)	1.8705	(0.07)	0.5782	1.3712	(0.04)	1.3437	(0.04)
200	Φ_1	0.1341	0.1263	(0.48)	0.1255	(0.51)	0.0775	0.0658	(0.75)	0.0659	(0.78)
	Φ_2	0.0797	0.0708	(0.84)	0.0703	(0.84)	0.0719	0.0645	(0.74)	0.0640	(0.77)
	Φ_3	0.4221	0.6484	(0.03)	0.6490	(0.02)	0.1328	0.1516	(0.30)	0.1483	(0.32)
	Φ_4	0.2151	0.2360	(0.35)	0.2334	(0.33)	0.1372	0.1598	(0.30)	0.1585	(0.29)
	Φ_5	1.3866	2.6349	(0.04)	2.6221	(0.04)	0.2316	0.5448	(0.04)	0.5402	(0.04)
	Φ_6	0.4919	1.0029	(0.17)	1.0041	(0.16)	0.2518	0.5919	(0.01)	0.5909	(0.02)
1000	Φ_1	0.0461	0.0566	(0.09)	0.0560	(0.09)	0.0126	0.0117	(0.57)	0.0114	(0.59)
	Φ_2	0.0249	0.0203	(0.88)	0.0201	(0.89)	0.0122	0.0117	(0.55)	0.0116	(0.57)
	Φ_3	0.2460	0.4086	(0.00)	0.4078	(0.00)	0.0258	0.0402	(0.10)	0.0399	(0.09)
	Φ_4	0.1049	0.0973	(0.65)	0.0968	(0.69)	0.0238	0.0376	(0.11)	0.0374	(0.11)
	Φ_5	0.6124	1.9030	(0.00)	1.8860	(0.00)	0.0480	0.1905	(0.00)	0.1877	(0.00)
	Φ_6	0.2715	0.4130	(0.23)	0.4006	(0.23)	0.0500	0.1706	(0.01)	0.1723	(0.01)

Table A1.4: MSE Table - Lasso FA, Model Error Involved

Variance of Parameter Estimates

Sample Size	Φ	Model 1					Model 2				
		ML	Lasso		Lasso (OD)		ML	Lasso		Lasso (OD)	
50	Φ_1	0.2259	0.2119	(0.67)	0.2047	(0.72)	0.2797	0.2411	(0.75)	0.2361	(0.82)
	Φ_2	0.2161	0.2053	(0.67)	0.2021	(0.69)	0.2136	0.1995	(0.70)	0.1958	(0.68)
	Φ_3	0.4939	0.5700	(0.33)	0.5451	(0.34)	0.6492	0.6247	(0.53)	0.6231	(0.55)
	Φ_4	0.5157	0.6889	(0.18)	0.6755	(0.19)	0.5825	0.7035	(0.25)	0.6898	(0.23)
	Φ_5	1.1269	2.3499	(0.07)	2.2159	(0.07)	6.1583	10.9808	(0.27)	11.3203	(0.28)
	Φ_6	2.1232	4.7051	(0.05)	4.4199	(0.05)	2.3465	3.1683	(0.04)	3.1252	(0.03)
100	Φ_1	0.1159	0.1067	(0.60)	0.1052	(0.65)	0.1314	0.1144	(0.76)	0.1102	(0.80)
	Φ_2	0.1008	0.0964	(0.55)	0.0944	(0.62)	0.1150	0.1055	(0.73)	0.1046	(0.79)
	Φ_3	0.2155	0.2482	(0.34)	0.2437	(0.30)	0.3020	0.3200	(0.45)	0.3091	(0.43)
	Φ_4	0.2271	0.2941	(0.29)	0.2882	(0.30)	0.2342	0.2934	(0.22)	0.2846	(0.23)
	Φ_5	0.5207	0.9886	(0.09)	0.9588	(0.07)	2.1160	1.0633	(0.25)	1.0429	(0.26)
	Φ_6	0.5663	1.4259	(0.01)	1.4005	(0.02)	0.6424	1.3766	(0.07)	1.3198	(0.07)
200	Φ_1	0.0544	0.0471	(0.63)	0.0465	(0.69)	0.0587	0.0472	(0.71)	0.0466	(0.67)
	Φ_2	0.0562	0.0542	(0.63)	0.0533	(0.69)	0.0578	0.0542	(0.68)	0.0531	(0.70)
	Φ_3	0.1035	0.1268	(0.25)	0.1241	(0.25)	0.1304	0.1372	(0.34)	0.1378	(0.35)
	Φ_4	0.0983	0.1210	(0.18)	0.1207	(0.18)	0.1215	0.1410	(0.29)	0.1389	(0.29)
	Φ_5	0.2461	0.4663	(0.08)	0.4569	(0.07)	0.3721	0.5839	(0.25)	0.5845	(0.25)
	Φ_6	0.2590	0.5848	(0.01)	0.5834	(0.01)	0.2793	0.6063	(0.02)	0.5984	(0.03)
1000	Φ_1	0.0114	0.0094	(0.61)	0.0092	(0.63)	0.0126	0.0103	(0.72)	0.0101	(0.71)
	Φ_2	0.0093	0.0088	(0.63)	0.0085	(0.68)	0.0104	0.0095	(0.66)	0.0094	(0.64)
	Φ_3	0.0210	0.0249	(0.22)	0.0249	(0.26)	0.0237	0.0247	(0.39)	0.0246	(0.42)
	Φ_4	0.0207	0.0255	(0.21)	0.0250	(0.21)	0.0210	0.0253	(0.24)	0.0251	(0.27)
	Φ_5	0.0416	0.0830	(0.03)	0.0830	(0.04)	0.0714	0.1101	(0.22)	0.1095	(0.21)
	Φ_6	0.0467	0.1078	(0.01)	0.1078	(0.01)	0.0496	0.1116	(0.04)	0.1113	(0.05)

Sample Size	Φ	Model 3					Model 4				
		ML	Lasso		Lasso (OD)		ML	Lasso		Lasso (OD)	
50	Φ_1	0.4715	0.3332	(0.90)	0.3317	(0.92)	0.3145	0.2710	(0.75)	0.2693	(0.76)
	Φ_2	0.3013	0.2611	(0.79)	0.2612	(0.79)	0.2576	0.2425	(0.73)	0.2388	(0.79)
	Φ_3	1.5027	1.0248	(0.69)	1.0177	(0.69)	0.5417	0.6092	(0.32)	0.5996	(0.35)
	Φ_4	0.7387	0.7865	(0.37)	0.7625	(0.38)	0.6411	0.7657	(0.19)	0.7352	(0.25)
	Φ_5	14.7520	14.0841	(0.48)	14.1723	(0.48)	1.5822	3.3832	(0.10)	2.6955	(0.07)
	Φ_6	5.6771	8.3943	(0.18)	8.6982	(0.17)	1.7714	3.7903	(0.05)	5.3037	(0.06)
100	Φ_1	0.2286	0.1682	(0.95)	0.1617	(0.97)	0.1479	0.1234	(0.78)	0.1229	(0.81)
	Φ_2	0.1314	0.1091	(0.89)	0.1078	(0.90)	0.1288	0.1172	(0.74)	0.1148	(0.82)
	Φ_3	0.6820	0.4507	(0.69)	0.4458	(0.69)	0.2551	0.2794	(0.39)	0.2738	(0.42)
	Φ_4	0.3287	0.3246	(0.48)	0.3264	(0.46)	0.2514	0.2883	(0.37)	0.2851	(0.40)
	Φ_5	42.3998	2.7022	(0.40)	2.6363	(0.42)	0.5202	0.9694	(0.06)	0.9716	(0.06)
	Φ_6	0.8856	1.4087	(0.22)	1.3544	(0.22)	0.6415	1.4946	(0.04)	1.4781	(0.04)
200	Φ_1	0.0969	0.0670	(0.91)	0.0659	(0.97)	0.0810	0.0669	(0.82)	0.0653	(0.88)
	Φ_2	0.0635	0.0543	(0.91)	0.0529	(0.91)	0.0741	0.0649	(0.80)	0.0644	(0.84)
	Φ_3	0.2875	0.2350	(0.66)	0.2293	(0.67)	0.1262	0.1353	(0.44)	0.1348	(0.48)
	Φ_4	0.1383	0.1414	(0.57)	0.1388	(0.60)	0.1346	0.1526	(0.30)	0.1488	(0.33)
	Φ_5	1.4418	1.6577	(0.38)	1.6604	(0.37)	0.2540	0.4991	(0.09)	0.4923	(0.08)
	Φ_6	0.3905	0.7652	(0.10)	0.7574	(0.09)	0.2561	0.5781	(0.05)	0.5711	(0.05)
1000	Φ_1	0.0192	0.0125	(0.95)	0.0123	(0.94)	0.0124	0.0103	(0.72)	0.0099	(0.84)
	Φ_2	0.0124	0.0096	(0.88)	0.0094	(0.91)	0.0139	0.0114	(0.79)	0.0112	(0.84)
	Φ_3	0.0538	0.0447	(0.70)	0.0441	(0.70)	0.0229	0.0229	(0.48)	0.0227	(0.52)
	Φ_4	0.0290	0.0272	(0.61)	0.0271	(0.62)	0.0240	0.0270	(0.36)	0.0268	(0.39)
	Φ_5	0.1598	0.3388	(0.42)	0.3445	(0.43)	0.0472	0.0939	(0.05)	0.0939	(0.05)
	Φ_6	0.0672	0.1173	(0.17)	0.1163	(0.19)	0.0499	0.1091	(0.03)	0.1073	(0.01)

Table A1.5: Variance Table - Lasso FA, No Model Error

Variance of Parameter Estimates

Sample Size	Φ	Model 1					Model 2				
		ML	Lasso		Lasso (OD)		ML	Lasso		Lasso (OD)	
50	Φ_1	0.2713	0.2462	(0.55)	0.2428	(0.57)	0.2584	0.2306	(0.65)	0.2225	(0.70)
	Φ_2	0.2326	0.2188	(0.65)	0.2175	(0.67)	0.2343	0.2243	(0.59)	0.2232	(0.61)
	Φ_3	0.4964	0.5572	(0.31)	0.5569	(0.31)	0.6829	0.5869	(0.52)	0.5701	(0.55)
	Φ_4	0.4838	0.6210	(0.18)	0.5925	(0.20)	0.6345	0.6978	(0.34)	0.6914	(0.32)
	Φ_5	1.1198	2.2136	(0.15)	2.2482	(0.16)	6.5494	3.7840	(0.29)	3.8707	(0.28)
	Φ_6	1.4877	2.7397	(0.07)	2.7240	(0.09)	5.0362	8.8095	(0.14)	8.0206	(0.15)
100	Φ_1	0.1070	0.0987	(0.57)	0.0952	(0.66)	0.1241	0.1099	(0.67)	0.1087	(0.69)
	Φ_2	0.1027	0.0980	(0.68)	0.0959	(0.76)	0.1099	0.0997	(0.67)	0.0985	(0.71)
	Φ_3	0.2248	0.2639	(0.34)	0.2528	(0.35)	0.2847	0.2582	(0.55)	0.2521	(0.54)
	Φ_4	0.2556	0.2916	(0.24)	0.2872	(0.25)	0.2423	0.2843	(0.27)	0.2808	(0.31)
	Φ_5	0.5835	0.9297	(0.12)	0.9243	(0.10)	2.8589	2.0384	(0.18)	2.0201	(0.17)
	Φ_6	0.5490	1.1915	(0.07)	1.1208	(0.06)	0.7196	1.5947	(0.05)	1.4736	(0.06)
200	Φ_1	0.0514	0.0463	(0.65)	0.0443	(0.72)	0.0688	0.0590	(0.72)	0.0581	(0.77)
	Φ_2	0.0558	0.0506	(0.63)	0.0497	(0.74)	0.0514	0.0478	(0.63)	0.0468	(0.71)
	Φ_3	0.1006	0.1081	(0.35)	0.1063	(0.41)	0.1517	0.1532	(0.36)	0.1520	(0.37)
	Φ_4	0.1172	0.1335	(0.25)	0.1319	(0.26)	0.1216	0.1470	(0.16)	0.1451	(0.19)
	Φ_5	0.2173	0.4077	(0.11)	0.3980	(0.12)	0.3804	1.3761	(0.07)	1.3916	(0.08)
	Φ_6	0.2893	0.5712	(0.09)	0.5657	(0.06)	0.3233	0.7566	(0.07)	0.7301	(0.07)
1000	Φ_1	0.0107	0.0091	(0.68)	0.0088	(0.72)	0.0112	0.0095	(0.72)	0.0093	(0.78)
	Φ_2	0.0101	0.0089	(0.66)	0.0088	(0.69)	0.0117	0.0104	(0.72)	0.0103	(0.70)
	Φ_3	0.0218	0.0234	(0.37)	0.0230	(0.35)	0.0257	0.0254	(0.52)	0.0249	(0.52)
	Φ_4	0.0194	0.0232	(0.19)	0.0230	(0.18)	0.0204	0.0248	(0.11)	0.0247	(0.15)
	Φ_5	0.0430	0.0771	(0.12)	0.0773	(0.12)	0.0657	1.0444	(0.00)	1.0543	(0.00)
	Φ_6	0.0461	0.0953	(0.05)	0.0946	(0.05)	0.0551	0.1210	(0.06)	0.1210	(0.04)

Sample Size	Φ	Model 3					Model 4				
		ML	Lasso		Lasso (OD)		ML	Lasso		Lasso (OD)	
50	Φ_1	0.4626	0.3482	(0.69)	0.3454	(0.70)	0.2888	0.2589	(0.71)	0.2560	(0.71)
	Φ_2	0.2837	0.2683	(0.66)	0.2642	(0.68)	0.2749	0.2520	(0.70)	0.2494	(0.70)
	Φ_3	1.1254	0.8523	(0.60)	0.8281	(0.61)	0.5328	0.5764	(0.38)	0.5795	(0.42)
	Φ_4	0.7080	0.7774	(0.39)	0.7641	(0.36)	0.6107	0.8206	(0.28)	0.7652	(0.27)
	Φ_5	161.4259	54.0144	(0.29)	53.7630	(0.29)	1.2073	2.4273	(0.12)	2.3815	(0.13)
	Φ_6	2.9347	7.1922	(0.14)	7.2000	(0.16)	3.0585	271.3923	(0.12)	271.6321	(0.12)
100	Φ_1	0.1988	0.1465	(0.77)	0.1500	(0.79)	0.1572	0.1380	(0.77)	0.1363	(0.80)
	Φ_2	0.1316	0.1189	(0.67)	0.1180	(0.66)	0.1488	0.1399	(0.73)	0.1384	(0.77)
	Φ_3	0.4662	0.4461	(0.52)	0.4247	(0.53)	0.2629	0.2816	(0.48)	0.2798	(0.45)
	Φ_4	0.3317	0.3816	(0.35)	0.3730	(0.33)	0.2618	0.2947	(0.32)	0.2931	(0.35)
	Φ_5	2.8333	3.6418	(0.31)	3.6026	(0.29)	0.5064	0.8303	(0.12)	0.8181	(0.12)
	Φ_6	0.8424	1.6278	(0.13)	1.5807	(0.12)	0.5782	1.3103	(0.03)	1.2826	(0.04)
200	Φ_1	0.1032	0.0813	(0.79)	0.0809	(0.81)	0.0775	0.0665	(0.82)	0.0665	(0.84)
	Φ_2	0.0670	0.0616	(0.69)	0.0613	(0.72)	0.0719	0.0644	(0.77)	0.0639	(0.81)
	Φ_3	0.2108	0.2387	(0.41)	0.2386	(0.38)	0.1328	0.1382	(0.57)	0.1352	(0.58)
	Φ_4	0.1349	0.1594	(0.28)	0.1566	(0.27)	0.1372	0.1498	(0.36)	0.1483	(0.39)
	Φ_5	0.9259	2.8361	(0.24)	2.8120	(0.23)	0.2316	0.4405	(0.13)	0.4375	(0.12)
	Φ_6	0.3022	0.7397	(0.03)	0.7411	(0.04)	0.2518	0.5229	(0.05)	0.5212	(0.05)
1000	Φ_1	0.0178	0.0147	(0.75)	0.0144	(0.77)	0.0126	0.0102	(0.83)	0.0099	(0.91)
	Φ_2	0.0133	0.0118	(0.65)	0.0116	(0.73)	0.0122	0.0105	(0.83)	0.0103	(0.88)
	Φ_3	0.0366	0.0432	(0.32)	0.0423	(0.37)	0.0258	0.0262	(0.46)	0.0260	(0.50)
	Φ_4	0.0294	0.0362	(0.14)	0.0361	(0.16)	0.0238	0.0255	(0.33)	0.0254	(0.37)
	Φ_5	0.0900	1.2798	(0.11)	1.2930	(0.10)	0.0480	0.0815	(0.15)	0.0808	(0.17)
	Φ_6	0.0621	0.1664	(0.02)	0.1549	(0.02)	0.0500	0.0975	(0.07)	0.0982	(0.06)

Table A1.6: Variance Table - Lasso FA, Model Error Involved

DE : Discrepancy due to Estimation (OLS)

Sample Size	Φ	Model 1					Model 2				
		ML	Lasso		Lasso (OD)		ML	Lasso		Lasso (OD)	
50	Φ_1	0.8328	0.5696	(0.84)	0.5342	(0.95)	0.7906	0.5522	(0.84)	0.5240	(0.98)
	Φ_2	0.7107	0.5227	(0.86)	0.4996	(0.97)	0.7024	0.4780	(0.88)	0.4626	(0.95)
	Φ_3	0.6453	0.4428	(0.84)	0.4229	(0.89)	0.6985	0.4397	(0.77)	0.4142	(0.88)
	Φ_4	0.5769	0.4289	(0.73)	0.4117	(0.79)	0.5503	0.4277	(0.69)	0.4041	(0.81)
	Φ_5	0.5304	0.4675	(0.51)	0.4423	(0.57)	0.5253	0.4900	(0.42)	0.4692	(0.49)
	Φ_6	0.5356	0.5368	(0.42)	0.5180	(0.46)	0.5047	0.5232	(0.40)	0.5081	(0.43)
100	Φ_1	0.4951	0.2976	(0.86)	0.2784	(0.98)	0.4820	0.3169	(0.74)	0.2902	(0.97)
	Φ_2	0.4397	0.3054	(0.84)	0.2937	(0.96)	0.4076	0.2819	(0.90)	0.2706	(0.95)
	Φ_3	0.4454	0.2816	(0.81)	0.2618	(0.87)	0.4138	0.2832	(0.76)	0.2662	(0.92)
	Φ_4	0.3437	0.2608	(0.65)	0.2484	(0.71)	0.3698	0.2694	(0.74)	0.2549	(0.86)
	Φ_5	0.3313	0.2873	(0.55)	0.2800	(0.62)	0.3533	0.3014	(0.41)	0.2868	(0.49)
	Φ_6	0.2961	0.3017	(0.41)	0.2950	(0.42)	0.2729	0.2904	(0.35)	0.2835	(0.36)
200	Φ_1	0.2761	0.1456	(0.83)	0.1332	(0.98)	0.2484	0.1448	(0.91)	0.1360	(0.99)
	Φ_2	0.2090	0.1374	(0.78)	0.1291	(0.95)	0.1870	0.1354	(0.76)	0.1262	(0.93)
	Φ_3	0.2008	0.1436	(0.70)	0.1322	(0.83)	0.1936	0.1403	(0.68)	0.1290	(0.81)
	Φ_4	0.1588	0.1224	(0.76)	0.1148	(0.85)	0.1824	0.1328	(0.71)	0.1271	(0.78)
	Φ_5	0.1585	0.1426	(0.51)	0.1383	(0.53)	0.1604	0.1534	(0.44)	0.1478	(0.52)
	Φ_6	0.1467	0.1510	(0.34)	0.1473	(0.42)	0.1432	0.1603	(0.30)	0.1569	(0.32)
1000	Φ_1	0.1243	0.0681	(0.84)	0.0624	(0.99)	0.1403	0.0796	(0.82)	0.0744	(0.95)
	Φ_2	0.1021	0.0699	(0.77)	0.0659	(0.93)	0.1020	0.0690	(0.75)	0.0650	(0.88)
	Φ_3	0.0962	0.0672	(0.73)	0.0629	(0.83)	0.1048	0.0648	(0.81)	0.0601	(0.94)
	Φ_4	0.0856	0.0638	(0.72)	0.0603	(0.84)	0.0890	0.0681	(0.78)	0.0657	(0.82)
	Φ_5	0.0852	0.0745	(0.53)	0.0712	(0.59)	0.0871	0.0789	(0.56)	0.0758	(0.59)
	Φ_6	0.0787	0.0786	(0.40)	0.0770	(0.41)	0.0739	0.0739	(0.36)	0.0719	(0.40)

Sample Size	Φ	Model 3				
		ML	Lasso		Lasso (OD)	
50	Φ_1	0.8137	0.6029	(0.84)	0.5651	(0.97)
	Φ_2	0.7710	0.5883	(0.87)	0.5661	(0.96)
	Φ_3	0.6857	0.5798	(0.79)	0.5522	(0.90)
	Φ_4	0.6248	0.5520	(0.78)	0.5318	(0.86)
	Φ_5	0.6025	0.6246	(0.51)	0.6052	(0.56)
	Φ_6	0.5721	0.6654	(0.45)	0.6554	(0.45)
100	Φ_1	0.4689	0.3544	(0.86)	0.3363	(0.96)
	Φ_2	0.4624	0.3588	(0.88)	0.3523	(0.96)
	Φ_3	0.4129	0.3698	(0.77)	0.3521	(0.86)
	Φ_4	0.3660	0.3604	(0.80)	0.3413	(0.86)
	Φ_5	0.3667	0.4261	(0.63)	0.4063	(0.65)
	Φ_6	0.3210	0.4318	(0.35)	0.4312	(0.41)
200	Φ_1	0.2531	0.1915	(0.85)	0.1760	(0.96)
	Φ_2	0.2309	0.1771	(0.85)	0.1688	(0.97)
	Φ_3	0.1971	0.2379	(0.74)	0.2139	(0.90)
	Φ_4	0.1926	0.2145	(0.86)	0.2084	(0.91)
	Φ_5	0.1738	0.2682	(0.49)	0.2572	(0.52)
	Φ_6	0.1546	0.2642	(0.40)	0.2585	(0.43)
1000	Φ_1	0.1274	0.1033	(0.90)	0.0976	(0.99)
	Φ_2	0.1076	0.1003	(0.86)	0.0945	(0.96)
	Φ_3	0.1016	0.1499	(0.86)	0.1436	(0.93)
	Φ_4	0.0920	0.1367	(0.80)	0.1325	(0.90)
	Φ_5	0.0839	0.1703	(0.57)	0.1655	(0.61)
	Φ_6	0.0799	0.1745	(0.46)	0.1704	(0.51)

Table A1.7: DE Table - Lasso SEM, No Model Error

DE : Discrepancy due to Estimation (OLS)

Sample Size	Φ	Model 1					Model 2				
		ML	Lasso		Lasso (OD)		ML	Lasso		Lasso (OD)	
50	Φ_1	0.7382	0.5604	(0.80)	0.5265	(0.89)	0.7990	0.6036	(0.80)	0.5690	(0.91)
	Φ_2	0.7328	0.5393	(0.84)	0.5239	(0.91)	0.7283	0.5784	(0.86)	0.5581	(0.92)
	Φ_3	0.6663	0.4609	(0.77)	0.4362	(0.81)	0.7769	0.5154	(0.79)	0.4761	(0.88)
	Φ_4	0.5757	0.4680	(0.67)	0.4392	(0.78)	0.5587	0.4565	(0.72)	0.4450	(0.78)
	Φ_5	0.5540	0.5097	(0.41)	0.4855	(0.45)	0.5861	0.5442	(0.44)	0.5252	(0.49)
	Φ_6	0.4986	0.5231	(0.31)	0.4983	(0.42)	0.5147	0.5330	(0.38)	0.5156	(0.41)
100	Φ_1	0.5033	0.3754	(0.76)	0.3481	(0.92)	0.5128	0.3693	(0.84)	0.3515	(0.91)
	Φ_2	0.4494	0.3416	(0.73)	0.3246	(0.86)	0.4376	0.3115	(0.80)	0.2971	(0.91)
	Φ_3	0.4658	0.2946	(0.77)	0.2791	(0.81)	0.3547	0.2735	(0.66)	0.2535	(0.80)
	Φ_4	0.3709	0.2757	(0.63)	0.2573	(0.73)	0.3725	0.2818	(0.62)	0.2709	(0.76)
	Φ_5	0.3366	0.3111	(0.43)	0.2914	(0.50)	0.3297	0.3255	(0.46)	0.3034	(0.52)
	Φ_6	0.3311	0.3046	(0.41)	0.2958	(0.43)	0.3277	0.3237	(0.34)	0.3146	(0.39)
200	Φ_1	0.2197	0.1489	(0.67)	0.1353	(0.89)	0.2141	0.1469	(0.80)	0.1365	(0.89)
	Φ_2	0.1923	0.1383	(0.74)	0.1310	(0.83)	0.2133	0.1521	(0.80)	0.1449	(0.86)
	Φ_3	0.1832	0.1352	(0.71)	0.1223	(0.85)	0.2041	0.1425	(0.77)	0.1350	(0.86)
	Φ_4	0.1675	0.1371	(0.60)	0.1269	(0.70)	0.1748	0.1359	(0.67)	0.1286	(0.75)
	Φ_5	0.1587	0.1577	(0.39)	0.1437	(0.45)	0.1721	0.1611	(0.45)	0.1542	(0.51)
	Φ_6	0.1427	0.1554	(0.35)	0.1505	(0.37)	0.1487	0.1640	(0.30)	0.1581	(0.35)
1000	Φ_1	0.1199	0.0736	(0.80)	0.0675	(0.91)	0.1082	0.0748	(0.76)	0.0676	(0.89)
	Φ_2	0.0975	0.0731	(0.76)	0.0689	(0.80)	0.0955	0.0687	(0.69)	0.0650	(0.84)
	Φ_3	0.1084	0.0702	(0.72)	0.0673	(0.85)	0.0941	0.0734	(0.68)	0.0671	(0.84)
	Φ_4	0.0879	0.0690	(0.56)	0.0651	(0.59)	0.0831	0.0661	(0.64)	0.0626	(0.73)
	Φ_5	0.0792	0.0753	(0.44)	0.0729	(0.52)	0.0895	0.0847	(0.48)	0.0823	(0.45)
	Φ_6	0.0708	0.0817	(0.26)	0.0795	(0.25)	0.0800	0.0874	(0.32)	0.0846	(0.38)

Sample Size	Φ	Model 3				
		ML	Lasso		Lasso (OD)	
50	Φ_1	0.8738	0.5700	(0.81)	0.5385	(0.93)
	Φ_2	0.7840	0.6067	(0.81)	0.5929	(0.91)
	Φ_3	0.7511	0.5096	(0.71)	0.4855	(0.84)
	Φ_4	0.6958	0.5163	(0.82)	0.5036	(0.88)
	Φ_5	0.6367	0.5536	(0.55)	0.5317	(0.66)
	Φ_6	0.5207	0.5503	(0.36)	0.5322	(0.38)
100	Φ_1	0.5444	0.3401	(0.88)	0.3156	(0.98)
	Φ_2	0.4162	0.3013	(0.86)	0.2866	(0.92)
	Φ_3	0.3731	0.2930	(0.71)	0.2654	(0.83)
	Φ_4	0.3482	0.2857	(0.74)	0.2758	(0.82)
	Φ_5	0.3578	0.3483	(0.45)	0.3326	(0.51)
	Φ_6	0.3543	0.3799	(0.33)	0.3697	(0.38)
200	Φ_1	0.2483	0.1536	(0.87)	0.1426	(0.98)
	Φ_2	0.2371	0.1620	(0.79)	0.1529	(0.94)
	Φ_3	0.1977	0.1383	(0.74)	0.1319	(0.80)
	Φ_4	0.2100	0.1443	(0.79)	0.1370	(0.84)
	Φ_5	0.1808	0.1715	(0.47)	0.1650	(0.52)
	Φ_6	0.1652	0.1779	(0.36)	0.1741	(0.41)
1000	Φ_1	0.1345	0.0724	(0.85)	0.0657	(0.98)
	Φ_2	0.1007	0.0743	(0.79)	0.0701	(0.89)
	Φ_3	0.1014	0.0758	(0.68)	0.0694	(0.78)
	Φ_4	0.1015	0.0723	(0.73)	0.0690	(0.82)
	Φ_5	0.0872	0.0831	(0.43)	0.0787	(0.45)
	Φ_6	0.0817	0.0849	(0.32)	0.0827	(0.40)

Table A1.8: DE Table - Lasso SEM, Model Error Involved

MSE : Mean Squared Error

Sample Size	Φ	Model 1					Model 2				
		ML	Lasso		Lasso (OD)		ML	Lasso		Lasso (OD)	
50	Φ_1	0.2001	0.2052	(0.42)	0.2029	(0.44)	0.2069	0.2099	(0.49)	0.2068	(0.51)
	Φ_2	0.1997	0.2123	(0.42)	0.2105	(0.40)	0.2224	0.2308	(0.47)	0.2304	(0.46)
	Φ_3	0.3195	0.3572	(0.33)	0.3572	(0.34)	0.3693	0.3900	(0.37)	0.3849	(0.37)
	Φ_4	0.3279	0.4049	(0.19)	0.4017	(0.17)	0.3614	0.4232	(0.18)	0.4231	(0.16)
	Φ_5	0.5152	0.7126	(0.13)	0.7118	(0.14)	0.8556	0.8332	(0.23)	0.8200	(0.18)
	Φ_6	0.6873	1.1628	(0.05)	1.1664	(0.04)	0.8029	1.0867	(0.14)	1.1318	(0.13)
100	Φ_1	0.1156	0.1180	(0.47)	0.1172	(0.45)	0.1304	0.1413	(0.41)	0.1389	(0.42)
	Φ_2	0.1187	0.1301	(0.32)	0.1287	(0.32)	0.1348	0.1414	(0.38)	0.1395	(0.40)
	Φ_3	0.1672	0.2097	(0.19)	0.2077	(0.17)	0.2206	0.2421	(0.33)	0.2398	(0.34)
	Φ_4	0.1910	0.2547	(0.04)	0.2543	(0.05)	0.2277	0.2798	(0.19)	0.2749	(0.19)
	Φ_5	0.3224	0.4898	(0.07)	0.4903	(0.09)	0.3936	0.5218	(0.16)	0.5235	(0.17)
	Φ_6	0.3524	0.6409	(0.02)	0.6494	(0.02)	0.3803	0.7020	(0.06)	0.7055	(0.06)
200	Φ_1	0.0570	0.0640	(0.30)	0.0634	(0.31)	0.0644	0.0684	(0.43)	0.0677	(0.44)
	Φ_2	0.0577	0.0626	(0.37)	0.0624	(0.36)	0.0635	0.0703	(0.31)	0.0702	(0.29)
	Φ_3	0.0867	0.1156	(0.18)	0.1148	(0.18)	0.1031	0.1326	(0.16)	0.1315	(0.18)
	Φ_4	0.0919	0.1295	(0.06)	0.1284	(0.08)	0.1055	0.1433	(0.10)	0.1428	(0.11)
	Φ_5	0.1479	0.2558	(0.02)	0.2559	(0.01)	0.1971	0.2985	(0.12)	0.2970	(0.11)
	Φ_6	0.1675	0.3512	(0.01)	0.3471	(0.01)	0.1932	0.4058	(0.02)	0.4107	(0.01)
1000	Φ_1	0.0267	0.0306	(0.37)	0.0304	(0.36)	0.0333	0.0383	(0.32)	0.0379	(0.34)
	Φ_2	0.0281	0.0342	(0.20)	0.0340	(0.21)	0.0330	0.0408	(0.23)	0.0404	(0.25)
	Φ_3	0.0429	0.0647	(0.09)	0.0644	(0.11)	0.0471	0.0674	(0.14)	0.0666	(0.14)
	Φ_4	0.0476	0.0797	(0.01)	0.0795	(0.01)	0.0517	0.0808	(0.06)	0.0807	(0.07)
	Φ_5	0.0720	0.1451	(0.01)	0.1454	(0.01)	0.0923	0.1662	(0.05)	0.1655	(0.05)
	Φ_6	0.0829	0.2378	(0.00)	0.2378	(0.00)	0.0869	0.2249	(0.00)	0.2234	(0.00)

Sample Size	Φ	Model 3				
		ML	Lasso		Lasso (OD)	
50	Φ_1	0.2927	0.2734	(0.61)	0.2710	(0.66)
	Φ_2	0.2761	0.2688	(0.58)	0.2677	(0.61)
	Φ_3	0.5351	0.4508	(0.55)	0.4497	(0.57)
	Φ_4	0.4241	0.4568	(0.33)	0.4542	(0.37)
	Φ_5	1.6341	1.0750	(0.40)	1.0884	(0.40)
	Φ_6	0.8703	1.2408	(0.13)	1.1865	(0.15)
100	Φ_1	0.1690	0.1531	(0.63)	0.1522	(0.61)
	Φ_2	0.1652	0.1600	(0.52)	0.1595	(0.54)
	Φ_3	0.3003	0.2867	(0.54)	0.2862	(0.52)
	Φ_4	0.2515	0.2805	(0.34)	0.2739	(0.35)
	Φ_5	0.7730	0.7489	(0.39)	0.7369	(0.39)
	Φ_6	0.4936	0.7937	(0.11)	0.8014	(0.11)
200	Φ_1	0.0788	0.0742	(0.59)	0.0738	(0.58)
	Φ_2	0.0765	0.0811	(0.44)	0.0802	(0.45)
	Φ_3	0.1467	0.1466	(0.47)	0.1451	(0.47)
	Φ_4	0.1279	0.1550	(0.18)	0.1541	(0.18)
	Φ_5	0.3033	0.3542	(0.23)	0.3503	(0.23)
	Φ_6	0.2136	0.4108	(0.03)	0.4125	(0.03)
1000	Φ_1	0.0410	0.0408	(0.51)	0.0407	(0.50)
	Φ_2	0.0379	0.0420	(0.37)	0.0412	(0.40)
	Φ_3	0.0636	0.0824	(0.20)	0.0823	(0.18)
	Φ_4	0.0545	0.0846	(0.09)	0.0839	(0.09)
	Φ_5	0.1377	0.1856	(0.22)	0.1842	(0.21)
	Φ_6	0.1031	0.2405	(0.01)	0.2384	(0.01)

Table A1.9: MSE Table - Lasso SEM, No Model Error

MSE : Mean Squared Error

Sample Size	Φ	Model 1					Model 2				
		ML	Lasso		Lasso (OD)		ML	Lasso		Lasso (OD)	
50	Φ_1	0.1912	0.1895	(0.59)	0.1880	(0.58)	0.2213	0.2249	(0.47)	0.2214	(0.47)
	Φ_2	0.1939	0.2027	(0.42)	0.2012	(0.43)	0.2417	0.2621	(0.39)	0.2595	(0.39)
	Φ_3	0.2909	0.3235	(0.32)	0.3217	(0.30)	0.3578	0.4048	(0.29)	0.4008	(0.32)
	Φ_4	0.3296	0.3785	(0.19)	0.3748	(0.22)	0.3788	0.4368	(0.16)	0.4332	(0.17)
	Φ_5	0.5367	0.7179	(0.11)	0.7205	(0.11)	2.8940	0.9103	(0.25)	0.8981	(0.25)
	Φ_6	0.6565	1.0077	(0.10)	1.0112	(0.08)	0.7149	1.0253	(0.10)	1.0187	(0.08)
100	Φ_1	0.1106	0.1115	(0.47)	0.1105	(0.45)	0.1318	0.1353	(0.50)	0.1346	(0.51)
	Φ_2	0.1213	0.1340	(0.37)	0.1323	(0.39)	0.1400	0.1533	(0.34)	0.1525	(0.32)
	Φ_3	0.1829	0.2063	(0.29)	0.2063	(0.27)	0.2209	0.2488	(0.27)	0.2473	(0.27)
	Φ_4	0.1959	0.2337	(0.22)	0.2317	(0.18)	0.2331	0.2919	(0.10)	0.2896	(0.12)
	Φ_5	0.2979	0.4436	(0.04)	0.4480	(0.04)	0.4096	0.5555	(0.16)	0.5524	(0.19)
	Φ_6	0.3252	0.5820	(0.04)	0.6134	(0.03)	0.3936	0.6451	(0.04)	0.6469	(0.04)
200	Φ_1	0.0508	0.0564	(0.32)	0.0563	(0.34)	0.0641	0.0693	(0.32)	0.0683	(0.38)
	Φ_2	0.0566	0.0647	(0.28)	0.0642	(0.28)	0.0712	0.0804	(0.29)	0.0799	(0.32)
	Φ_3	0.0809	0.1004	(0.19)	0.0998	(0.19)	0.1172	0.1463	(0.09)	0.1457	(0.11)
	Φ_4	0.0929	0.1241	(0.12)	0.1241	(0.11)	0.1180	0.1487	(0.05)	0.1490	(0.05)
	Φ_5	0.1432	0.2321	(0.09)	0.2305	(0.07)	0.2328	0.3182	(0.15)	0.3156	(0.15)
	Φ_6	0.1637	0.3497	(0.00)	0.3484	(0.00)	0.2157	0.4168	(0.01)	0.4168	(0.01)
1000	Φ_1	0.0270	0.0332	(0.26)	0.0328	(0.27)	0.0342	0.0419	(0.23)	0.0414	(0.23)
	Φ_2	0.0282	0.0343	(0.25)	0.0340	(0.24)	0.0441	0.0510	(0.24)	0.0508	(0.25)
	Φ_3	0.0430	0.0605	(0.13)	0.0601	(0.13)	0.0672	0.0895	(0.13)	0.0895	(0.13)
	Φ_4	0.0437	0.0694	(0.04)	0.0694	(0.05)	0.0767	0.1068	(0.05)	0.1059	(0.04)
	Φ_5	0.0730	0.1397	(0.00)	0.1389	(0.00)	0.1321	0.2156	(0.02)	0.2149	(0.03)
	Φ_6	0.0798	0.1981	(0.02)	0.1992	(0.02)	0.1278	0.2522	(0.00)	0.2525	(0.00)

Sample Size	Φ	Model 3				
		ML	Lasso		Lasso (OD)	
50	Φ_1	0.3228	0.3253	(0.54)	0.3220	(0.56)
	Φ_2	0.2966	0.3020	(0.44)	0.2998	(0.48)
	Φ_3	0.6500	0.6728	(0.43)	0.6707	(0.43)
	Φ_4	0.5502	0.5830	(0.38)	0.5805	(0.38)
	Φ_5	1.4499	1.3556	(0.31)	1.3458	(0.34)
	Φ_6	0.9430	1.2889	(0.17)	1.2241	(0.16)
100	Φ_1	0.2335	0.2325	(0.51)	0.2311	(0.52)
	Φ_2	0.1877	0.1933	(0.43)	0.1938	(0.41)
	Φ_3	0.4482	0.4810	(0.25)	0.4792	(0.26)
	Φ_4	0.3308	0.3624	(0.20)	0.3618	(0.20)
	Φ_5	1.4961	0.9840	(0.29)	0.9909	(0.26)
	Φ_6	0.6687	0.9275	(0.06)	0.9275	(0.07)
200	Φ_1	0.1360	0.1426	(0.34)	0.1424	(0.32)
	Φ_2	0.1100	0.1142	(0.45)	0.1136	(0.44)
	Φ_3	0.3063	0.3371	(0.13)	0.3368	(0.15)
	Φ_4	0.2252	0.2492	(0.21)	0.2489	(0.18)
	Φ_5	0.6680	0.7357	(0.23)	0.7277	(0.28)
	Φ_6	0.4198	0.5551	(0.11)	0.5552	(0.12)
1000	Φ_1	0.0940	0.1039	(0.22)	0.1036	(0.22)
	Φ_2	0.0728	0.0775	(0.31)	0.0772	(0.31)
	Φ_3	0.2375	0.2736	(0.03)	0.2735	(0.02)
	Φ_4	0.1530	0.1710	(0.18)	0.1708	(0.18)
	Φ_5	0.4979	0.5620	(0.16)	0.5620	(0.14)
	Φ_6	0.2884	0.3762	(0.09)	0.3756	(0.09)

Table A1.10: MSE Table - Lasso SEM, Model Error Involved

Variance of Parameter Estimates

Sample Size	Φ	Model 1					Model 2				
		ML	Lasso		Lasso (OD)		ML	Lasso		Lasso (OD)	
50	Φ_1	0.2001	0.2025	(0.50)	0.2005	(0.54)	0.2069	0.2045	(0.51)	0.2012	(0.55)
	Φ_2	0.1997	0.2087	(0.38)	0.2070	(0.43)	0.2224	0.2236	(0.44)	0.2230	(0.45)
	Φ_3	0.3195	0.3488	(0.27)	0.3482	(0.31)	0.3693	0.3615	(0.44)	0.3571	(0.45)
	Φ_4	0.3279	0.3915	(0.16)	0.3890	(0.15)	0.3614	0.4025	(0.23)	0.4017	(0.24)
	Φ_5	0.5152	0.6821	(0.12)	0.6814	(0.13)	0.8556	0.7858	(0.25)	0.7720	(0.20)
	Φ_6	0.6873	1.1087	(0.04)	1.1148	(0.03)	0.8029	1.0379	(0.16)	1.0813	(0.15)
100	Φ_1	0.1156	0.1152	(0.51)	0.1144	(0.53)	0.1304	0.1325	(0.48)	0.1302	(0.48)
	Φ_2	0.1187	0.1247	(0.32)	0.1234	(0.35)	0.1348	0.1339	(0.40)	0.1322	(0.43)
	Φ_3	0.1672	0.1872	(0.26)	0.1855	(0.24)	0.2206	0.2230	(0.33)	0.2203	(0.39)
	Φ_4	0.1910	0.2330	(0.06)	0.2328	(0.05)	0.2277	0.2591	(0.21)	0.2544	(0.20)
	Φ_5	0.3224	0.4574	(0.04)	0.4579	(0.04)	0.3896	0.4720	(0.20)	0.4747	(0.20)
	Φ_6	0.3524	0.5967	(0.02)	0.6050	(0.01)	0.3803	0.6485	(0.11)	0.6530	(0.10)
200	Φ_1	0.0570	0.0562	(0.49)	0.0556	(0.54)	0.0644	0.0603	(0.61)	0.0595	(0.58)
	Φ_2	0.0577	0.0593	(0.40)	0.0591	(0.37)	0.0635	0.0636	(0.42)	0.0635	(0.42)
	Φ_3	0.0867	0.0973	(0.18)	0.0965	(0.22)	0.1031	0.1119	(0.24)	0.1110	(0.28)
	Φ_4	0.0919	0.1125	(0.10)	0.1118	(0.11)	0.1055	0.1248	(0.13)	0.1243	(0.14)
	Φ_5	0.1479	0.2220	(0.05)	0.2219	(0.06)	0.1971	0.2518	(0.20)	0.2503	(0.19)
	Φ_6	0.1675	0.3103	(0.03)	0.3070	(0.03)	0.1932	0.3460	(0.04)	0.3506	(0.02)
1000	Φ_1	0.0267	0.0257	(0.53)	0.0256	(0.56)	0.0333	0.0325	(0.51)	0.0322	(0.54)
	Φ_2	0.0281	0.0293	(0.34)	0.0291	(0.35)	0.0330	0.0337	(0.35)	0.0334	(0.38)
	Φ_3	0.0429	0.0491	(0.15)	0.0489	(0.15)	0.0471	0.0482	(0.40)	0.0478	(0.39)
	Φ_4	0.0476	0.0596	(0.05)	0.0594	(0.05)	0.0517	0.0600	(0.17)	0.0599	(0.17)
	Φ_5	0.0720	0.1092	(0.00)	0.1095	(0.00)	0.0923	0.1198	(0.20)	0.1192	(0.20)
	Φ_6	0.0829	0.1756	(0.00)	0.1756	(0.00)	0.0869	0.1597	(0.03)	0.1584	(0.03)

Sample Size	Φ	Model 3				
		ML	Lasso		Lasso (OD)	
50	Φ_1	0.2927	0.2627	(0.68)	0.2608	(0.74)
	Φ_2	0.2761	0.2589	(0.68)	0.2578	(0.70)
	Φ_3	0.5350	0.4250	(0.70)	0.4234	(0.73)
	Φ_4	0.4241	0.4288	(0.42)	0.4272	(0.43)
	Φ_5	1.6014	0.9474	(0.54)	0.9655	(0.52)
	Φ_6	0.8703	1.2003	(0.18)	1.1441	(0.23)
100	Φ_1	0.1690	0.1471	(0.76)	0.1459	(0.80)
	Φ_2	0.1652	0.1501	(0.54)	0.1493	(0.58)
	Φ_3	0.3003	0.2584	(0.69)	0.2581	(0.71)
	Φ_4	0.2515	0.2541	(0.45)	0.2482	(0.47)
	Φ_5	0.7652	0.6486	(0.54)	0.6409	(0.53)
	Φ_6	0.4935	0.7329	(0.10)	0.7389	(0.10)
200	Φ_1	0.0788	0.0654	(0.82)	0.0652	(0.85)
	Φ_2	0.0765	0.0734	(0.54)	0.0726	(0.55)
	Φ_3	0.1467	0.1172	(0.76)	0.1163	(0.78)
	Φ_4	0.1279	0.1315	(0.36)	0.1307	(0.36)
	Φ_5	0.3033	0.2873	(0.38)	0.2840	(0.39)
	Φ_6	0.2136	0.3344	(0.11)	0.3362	(0.11)
1000	Φ_1	0.0410	0.0336	(0.81)	0.0338	(0.82)
	Φ_2	0.0379	0.0339	(0.65)	0.0332	(0.70)
	Φ_3	0.0636	0.0545	(0.75)	0.0545	(0.76)
	Φ_4	0.0545	0.0587	(0.37)	0.0583	(0.35)
	Φ_5	0.1377	0.1206	(0.57)	0.1200	(0.58)
	Φ_6	0.1031	0.1744	(0.09)	0.1729	(0.09)

Table A1.11: Variance Table - Lasso SEM, No Model Error

Variance of Parameter Estimates

Sample Size	Φ	Model 1					Model 2				
		ML	Lasso		Lasso (OD)		ML	Lasso		Lasso (OD)	
50	Φ_1	0.1912	0.1891	(0.57)	0.1878	(0.61)	0.2162	0.2133	(0.53)	0.2100	(0.53)
	Φ_2	0.1939	0.2009	(0.33)	0.1993	(0.38)	0.2297	0.2390	(0.35)	0.2366	(0.35)
	Φ_3	0.2909	0.3100	(0.28)	0.3083	(0.30)	0.3645	0.3744	(0.38)	0.3717	(0.41)
	Φ_4	0.3296	0.3641	(0.21)	0.3610	(0.24)	0.3602	0.3885	(0.25)	0.3860	(0.25)
	Φ_5	0.5367	0.7061	(0.13)	0.7094	(0.11)	2.9301	0.8344	(0.27)	0.8224	(0.29)
	Φ_6	0.6565	0.9759	(0.07)	0.9774	(0.08)	0.6738	0.9365	(0.12)	0.9290	(0.13)
100	Φ_1	0.1106	0.1103	(0.51)	0.1093	(0.52)	0.1278	0.1248	(0.62)	0.1243	(0.62)
	Φ_2	0.1213	0.1283	(0.36)	0.1267	(0.39)	0.1286	0.1336	(0.39)	0.1326	(0.50)
	Φ_3	0.1829	0.1948	(0.30)	0.1949	(0.32)	0.2075	0.2123	(0.38)	0.2104	(0.40)
	Φ_4	0.1959	0.2219	(0.20)	0.2205	(0.17)	0.2099	0.2442	(0.13)	0.2420	(0.15)
	Φ_5	0.2979	0.4185	(0.06)	0.4218	(0.03)	0.3589	0.4589	(0.21)	0.4517	(0.24)
	Φ_6	0.3252	0.5315	(0.07)	0.5642	(0.09)	0.3540	0.5477	(0.07)	0.5480	(0.06)
200	Φ_1	0.0508	0.0506	(0.47)	0.0505	(0.45)	0.0579	0.0566	(0.59)	0.0557	(0.61)
	Φ_2	0.0566	0.0589	(0.34)	0.0584	(0.35)	0.0600	0.0609	(0.48)	0.0607	(0.45)
	Φ_3	0.0809	0.0858	(0.30)	0.0856	(0.29)	0.1007	0.1064	(0.32)	0.1062	(0.29)
	Φ_4	0.0929	0.1099	(0.07)	0.1096	(0.11)	0.0982	0.1114	(0.12)	0.1113	(0.13)
	Φ_5	0.1432	0.2063	(0.03)	0.2054	(0.04)	0.1991	0.2357	(0.18)	0.2337	(0.18)
	Φ_6	0.1637	0.3056	(0.00)	0.3045	(0.00)	0.1836	0.3220	(0.02)	0.3207	(0.03)
1000	Φ_1	0.0270	0.0264	(0.51)	0.0261	(0.53)	0.0282	0.0281	(0.48)	0.0279	(0.49)
	Φ_2	0.0282	0.0289	(0.35)	0.0287	(0.37)	0.0316	0.0322	(0.38)	0.0321	(0.37)
	Φ_3	0.0430	0.0469	(0.19)	0.0466	(0.24)	0.0487	0.0520	(0.23)	0.0519	(0.24)
	Φ_4	0.0437	0.0530	(0.07)	0.0528	(0.08)	0.0474	0.0562	(0.05)	0.0559	(0.06)
	Φ_5	0.0730	0.1116	(0.03)	0.1110	(0.02)	0.0887	0.1276	(0.06)	0.1269	(0.06)
	Φ_6	0.0798	0.1528	(0.01)	0.1537	(0.01)	0.0939	0.1718	(0.00)	0.1715	(0.00)

Sample Size	Φ	Model 3				
		ML	Lasso		Lasso (OD)	
50	Φ_1	0.2736	0.2584	(0.56)	0.2551	(0.54)
	Φ_2	0.2593	0.2493	(0.57)	0.2471	(0.55)
	Φ_3	0.4832	0.4593	(0.54)	0.4551	(0.58)
	Φ_4	0.4317	0.4365	(0.37)	0.4348	(0.37)
	Φ_5	1.1988	0.9931	(0.43)	0.9757	(0.40)
	Φ_6	0.7366	1.0537	(0.18)	0.9899	(0.17)
100	Φ_1	0.1692	0.1530	(0.69)	0.1514	(0.75)
	Φ_2	0.1485	0.1433	(0.56)	0.1444	(0.57)
	Φ_3	0.2629	0.2539	(0.53)	0.2509	(0.53)
	Φ_4	0.2291	0.2414	(0.19)	0.2407	(0.19)
	Φ_5	1.1163	0.5520	(0.32)	0.5482	(0.31)
	Φ_6	0.4530	0.6834	(0.08)	0.6826	(0.08)
200	Φ_1	0.0773	0.0724	(0.67)	0.0727	(0.67)
	Φ_2	0.0756	0.0734	(0.55)	0.0729	(0.59)
	Φ_3	0.1278	0.1240	(0.51)	0.1237	(0.47)
	Φ_4	0.1161	0.1267	(0.24)	0.1265	(0.27)
	Φ_5	0.2445	0.2754	(0.31)	0.2697	(0.30)
	Φ_6	0.2056	0.3258	(0.05)	0.3253	(0.05)
1000	Φ_1	0.0369	0.0345	(0.67)	0.0343	(0.74)
	Φ_2	0.0377	0.0361	(0.51)	0.0359	(0.50)
	Φ_3	0.0597	0.0608	(0.35)	0.0608	(0.38)
	Φ_4	0.0561	0.0610	(0.24)	0.0611	(0.25)
	Φ_5	0.1092	0.1296	(0.25)	0.1294	(0.21)
	Φ_6	0.0954	0.1561	(0.00)	0.1556	(0.01)

Table A1.12: Variance Table - Lasso SEM, Model Error Involved

A2. Result Tables for BLasso

In this subsection, some result tables related to BLasso are provided. They include observed outcomes of OD, DE, sMSE, and Variance. Table A2.1∼A2.8 contain BLasso FA outcomes, while the others, A2.9∼A2.16 display BLasso SEM results. Cells with gray color indicates that Lasso yields smaller mean estimates than BLasso for the corresponding conditions to those cells. Some notable results can be summarized as follows.

- Lasso produces less DE over almost all the conditions. OD shows the similar trend, but not as much as DE. It seems that BLasso is able to yield less DA than Lasso.
- BLasso outperforms Lasso in producing smaller sMSE and Variance, especially in SEM. However, the number of case that BLasso yields less sMSE and variance than ML is no better than that of Lasso.
- In FA, Lasso outperforms BLasso in producing smaller sMSE and variance at almost the same condition that Lasso does with regard to ML.
- Overall, it seems that BLasso performs as an intermediate of ML and Lasso. This opinion can be supported by the fact that Bayesian estimation is asymptotically equal to MLE under some conditions.

OD : Overall Discrepancy (OLS)

Sample Size	Φ	Model 1			Model 2		
		ML	BLasso (MAP)	BLasso (Posterior Mean)	ML	BLasso (MAP)	BLasso (Posterior Mean)
50	Φ_1	0.9443	0.8535	1.1857	1.0040	0.8406	1.1308
	Φ_2	0.7676	0.6068	0.7651	0.7378	0.5944	0.7758
	Φ_3	0.7939	0.6699	0.8707	0.7790	0.6779	0.9355
	Φ_4	0.6764	0.6089	0.8887	0.6597	0.5451	0.7782
	Φ_5	0.6738	0.7276	1.0618	0.5733	0.6583	0.9709
	Φ_6	0.5401	0.6097	0.8505	0.5333	0.5736	0.8460
100	Φ_1	0.3873	0.3536	0.4035	0.4576	0.4264	0.4647
	Φ_2	0.3551	0.3338	0.3617	0.3765	0.3324	0.3758
	Φ_3	0.3950	0.3708	0.4703	0.3374	0.3103	0.4034
	Φ_4	0.3391	0.3162	0.3858	0.2972	0.2781	0.3636
	Φ_5	0.2872	0.3216	0.4001	0.3091	0.3674	0.4727
	Φ_6	0.2748	0.3150	0.3987	0.2889	0.3445	0.4272
200	Φ_1	0.2329	0.2340	0.2562	0.2222	0.2255	0.2412
	Φ_2	0.2090	0.2086	0.2073	0.2020	0.1960	0.1990
	Φ_3	0.1653	0.1615	0.1804	0.1860	0.1832	0.1943
	Φ_4	0.1315	0.1252	0.1350	0.1815	0.1743	0.1818
	Φ_5	0.1604	0.1741	0.1894	0.1584	0.1774	0.1919
	Φ_6	0.1326	0.1448	0.1627	0.1373	0.1553	0.1693
1000	Φ_1	0.0475	0.0501	0.0507	0.0440	0.0475	0.0474
	Φ_2	0.0316	0.0331	0.0342	0.0354	0.0379	0.0382
	Φ_3	0.0360	0.0360	0.0361	0.0307	0.0316	0.0298
	Φ_4	0.0310	0.0317	0.0305	0.0305	0.0319	0.0308
	Φ_5	0.0294	0.0312	0.0323	0.0327	0.0351	0.0356
	Φ_6	0.0254	0.0276	0.0270	0.0273	0.0290	0.0299

Sample Size	Φ	Model 3			Model 4		
		ML	BLasso (MAP)	BLasso (Posterior Mean)	ML	BLasso (MAP)	BLasso (Posterior Mean)
50	Φ_1	1.0567	0.9006	1.1466	1.0787	0.9627	1.2632
	Φ_2	0.8863	0.7602	0.9302	0.7471	0.6520	0.7928
	Φ_3	0.8288	0.7063	1.1054	0.9087	0.8185	1.1775
	Φ_4	0.6776	0.5782	0.8093	0.6764	0.6382	0.8135
	Φ_5	0.7372	0.8286	1.2672	0.6997	0.8147	1.0689
	Φ_6	0.6785	0.7601	1.0644	0.6566	0.7365	0.9974
100	Φ_1	0.4565	0.4420	0.4744	0.5584	0.5433	0.6304
	Φ_2	0.4251	0.3803	0.4568	0.4210	0.4280	0.4662
	Φ_3	0.3734	0.3496	0.3960	0.4489	0.4323	0.5090
	Φ_4	0.3696	0.3220	0.4114	0.3504	0.3434	0.4066
	Φ_5	0.3402	0.3846	0.4632	0.3438	0.4080	0.5040
	Φ_6	0.2973	0.3411	0.4247	0.3018	0.3462	0.4214
200	Φ_1	0.2379	0.2342	0.2487	0.3010	0.3032	0.3173
	Φ_2	0.2048	0.1950	0.2053	0.2213	0.2186	0.2421
	Φ_3	0.2109	0.2066	0.2281	0.2283	0.2318	0.2489
	Φ_4	0.1716	0.1693	0.1773	0.1933	0.1972	0.2087
	Φ_5	0.1739	0.1915	0.2096	0.1599	0.1902	0.2144
	Φ_6	0.1572	0.1716	0.1978	0.1677	0.1934	0.2210
1000	Φ_1	0.0449	0.0475	0.0448	0.0417	0.0446	0.0469
	Φ_2	0.0398	0.0414	0.0411	0.0460	0.0501	0.0496
	Φ_3	0.0363	0.0379	0.0370	0.0433	0.0443	0.0454
	Φ_4	0.0356	0.0364	0.0354	0.0354	0.0370	0.0361
	Φ_5	0.0334	0.0365	0.0363	0.0303	0.0336	0.0331
	Φ_6	0.0295	0.0310	0.0315	0.0306	0.0322	0.0330

Table A2.1: OD Table : BLasso FA, No Model Error

OD : Overall Discrepancy (OLS)

Sample Size	Φ	Model 1			Model 2		
		ML	BLasso (MAP)	BLasso (Posterior Mean)	ML	BLasso (MAP)	BLasso (Posterior Mean)
50	Φ_1	1.0726	1.2306	1.5961	0.9532	0.7939	1.1112
	Φ_2	0.8783	0.7809	1.0762	0.8639	0.6855	0.9094
	Φ_3	0.7632	0.6729	0.8833	0.7784	0.6768	0.9367
	Φ_4	0.6865	0.6116	0.8010	0.7361	0.5819	0.7991
	Φ_5	0.6796	0.7627	1.0746	0.6495	0.6859	0.9304
	Φ_6	0.6207	0.7030	0.9393	0.6611	0.6790	0.9220
100	Φ_1	0.4735	0.4208	0.4570	0.4719	0.4429	0.4720
	Φ_2	0.3907	0.3332	0.3598	0.4100	0.3707	0.4228
	Φ_3	0.3945	0.3688	0.4131	0.4091	0.3755	0.4498
	Φ_4	0.4256	0.3684	0.4181	0.3608	0.3134	0.3643
	Φ_5	0.3350	0.3637	0.4593	0.3514	0.3873	0.4757
	Φ_6	0.3188	0.3489	0.4006	0.3281	0.3582	0.4235
200	Φ_1	0.2282	0.2111	0.2301	0.2497	0.2340	0.2488
	Φ_2	0.2323	0.2094	0.2321	0.2093	0.1895	0.2022
	Φ_3	0.2135	0.2047	0.2192	0.2197	0.2061	0.2228
	Φ_4	0.2205	0.2053	0.2125	0.2110	0.1964	0.2154
	Φ_5	0.1727	0.2027	0.2185	0.1744	0.1906	0.2035
	Φ_6	0.1860	0.2043	0.2151	0.2154	0.2231	0.2508
1000	Φ_1	0.0549	0.0494	0.0487	0.0521	0.0451	0.0433
	Φ_2	0.0635	0.0519	0.0505	0.0600	0.0510	0.0503
	Φ_3	0.0625	0.0617	0.0607	0.0612	0.0613	0.0600
	Φ_4	0.0777	0.0771	0.0756	0.0783	0.0777	0.0761
	Φ_5	0.0636	0.0672	0.0668	0.0624	0.0673	0.0646
	Φ_6	0.0815	0.0856	0.0852	0.0851	0.0877	0.0877

Sample Size	Φ	Model 3			Model 4		
		ML	BLasso (MAP)	BLasso (Posterior Mean)	ML	BLasso (MAP)	BLasso (Posterior Mean)
50	Φ_1	0.9478	0.7860	1.0017	1.1193	0.9805	1.2857
	Φ_2	0.9701	0.7472	1.0156	0.8930	0.7643	0.9956
	Φ_3	0.7324	0.6506	0.8840	0.8984	0.7932	1.0108
	Φ_4	0.8066	0.6825	0.9444	0.7597	0.6839	0.9281
	Φ_5	0.6964	0.7590	1.0830	0.7423	0.8520	1.1077
	Φ_6	0.6446	0.6787	0.9850	0.7030	0.7803	1.0300
100	Φ_1	0.5624	0.5099	0.5679	0.5781	0.5652	0.6125
	Φ_2	0.3690	0.3241	0.3455	0.4554	0.4204	0.4631
	Φ_3	0.3730	0.3493	0.4134	0.4855	0.4683	0.5230
	Φ_4	0.3939	0.3433	0.4178	0.4205	0.4113	0.4795
	Φ_5	0.3263	0.3783	0.4814	0.3757	0.4457	0.5195
	Φ_6	0.3418	0.3725	0.4833	0.3840	0.4352	0.5096
200	Φ_1	0.2485	0.2438	0.2466	0.3069	0.3102	0.3147
	Φ_2	0.2329	0.2194	0.2209	0.2584	0.2462	0.2644
	Φ_3	0.2095	0.2075	0.2235	0.2320	0.2336	0.2401
	Φ_4	0.1964	0.1937	0.2019	0.2255	0.2232	0.2316
	Φ_5	0.1832	0.1976	0.2227	0.2088	0.2442	0.2712
	Φ_6	0.1605	0.1779	0.1981	0.2069	0.2393	0.2632
1000	Φ_1	0.0514	0.0501	0.0486	0.0553	0.0528	0.0519
	Φ_2	0.0471	0.0462	0.0462	0.0633	0.0521	0.0518
	Φ_3	0.0432	0.0432	0.0431	0.0664	0.0668	0.0647
	Φ_4	0.0518	0.0521	0.0509	0.0803	0.0795	0.0791
	Φ_5	0.0430	0.0463	0.0469	0.0672	0.0723	0.0715
	Φ_6	0.0413	0.0410	0.0412	0.0833	0.0869	0.0865

Table A2.2: OD Table : BLasso FA, Model Error Involved

DE : Discrepancy due to Estimation (OLS)

Sample Size	Φ	Model 1			Model 2		
		ML	BLasso (MAP)	BLasso (Posterior Mean)	ML	BLasso (MAP)	BLasso (Posterior Mean)
50	Φ_1	0.9443	0.8528	1.1823	1.0040	0.8400	1.1282
	Φ_2	0.7676	0.6074	0.7634	0.7377	0.5940	0.7741
	Φ_3	0.7939	0.6700	0.8703	0.7790	0.6782	0.9348
	Φ_4	0.6764	0.6082	0.8882	0.6597	0.5452	0.7781
	Φ_5	0.6738	0.7275	1.0609	0.5618	0.6447	0.9504
	Φ_6	0.5401	0.6100	0.8493	0.5226	0.5616	0.8279
100	Φ_1	0.3873	0.3531	0.4018	0.4576	0.4263	0.4634
	Φ_2	0.3550	0.3340	0.3608	0.3764	0.3322	0.3749
	Φ_3	0.3950	0.3708	0.4700	0.3374	0.3107	0.4030
	Φ_4	0.3390	0.3158	0.3855	0.2972	0.2779	0.3636
	Φ_5	0.2872	0.3213	0.3994	0.3091	0.3668	0.4719
	Φ_6	0.2748	0.3148	0.3976	0.2889	0.3436	0.4264
200	Φ_1	0.2329	0.2332	0.2547	0.2222	0.2247	0.2399
	Φ_2	0.2089	0.2088	0.2066	0.2020	0.1958	0.1986
	Φ_3	0.1653	0.1615	0.1803	0.1860	0.1833	0.1941
	Φ_4	0.1315	0.1252	0.1348	0.1815	0.1743	0.1817
	Φ_5	0.1604	0.1740	0.1890	0.1584	0.1772	0.1915
	Φ_6	0.1326	0.1444	0.1621	0.1373	0.1556	0.1689
1000	Φ_1	0.0475	0.0495	0.0500	0.0440	0.0471	0.0469
	Φ_2	0.0316	0.0328	0.0337	0.0354	0.0376	0.0377
	Φ_3	0.0360	0.0360	0.0360	0.0307	0.0316	0.0297
	Φ_4	0.0310	0.0318	0.0305	0.0305	0.0319	0.0308
	Φ_5	0.0294	0.0312	0.0322	0.0327	0.0351	0.0355
	Φ_6	0.0254	0.0274	0.0269	0.0273	0.0290	0.0297

Sample Size	Φ	Model 3			Model 4		
		ML	BLasso (MAP)	BLasso (Posterior Mean)	ML	BLasso (MAP)	BLasso (Posterior Mean)
50	Φ_1	1.0567	0.8996	1.1440	1.0787	0.9617	1.2600
	Φ_2	0.8863	0.7608	0.9291	0.7471	0.6527	0.7912
	Φ_3	0.8039	0.6848	1.0719	0.9087	0.8181	1.1770
	Φ_4	0.6776	0.5783	0.8092	0.6764	0.6367	0.8128
	Φ_5	0.6340	0.7119	1.0883	0.6996	0.8125	1.0679
	Φ_6	0.6581	0.7373	1.0320	0.6566	0.7353	0.9959
100	Φ_1	0.4565	0.4415	0.4732	0.5584	0.5424	0.6283
	Φ_2	0.4251	0.3801	0.4558	0.4210	0.4281	0.4650
	Φ_3	0.3734	0.3492	0.3958	0.4489	0.4313	0.5087
	Φ_4	0.3696	0.3221	0.4114	0.3504	0.3428	0.4061
	Φ_5	0.3164	0.3569	0.4299	0.3438	0.4070	0.5033
	Φ_6	0.2973	0.3412	0.4244	0.3018	0.3455	0.4205
200	Φ_1	0.2379	0.2337	0.2478	0.3010	0.3027	0.3160
	Φ_2	0.2048	0.1947	0.2047	0.2213	0.2176	0.2411
	Φ_3	0.2109	0.2063	0.2280	0.2283	0.2312	0.2487
	Φ_4	0.1716	0.1697	0.1773	0.1933	0.1967	0.2084
	Φ_5	0.1722	0.1889	0.2069	0.1599	0.1897	0.2140
	Φ_6	0.1572	0.1716	0.1976	0.1677	0.1929	0.2203
1000	Φ_1	0.0449	0.0473	0.0445	0.0417	0.0442	0.0462
	Φ_2	0.0398	0.0411	0.0408	0.0460	0.0494	0.0491
	Φ_3	0.0363	0.0379	0.0370	0.0433	0.0444	0.0453
	Φ_4	0.0356	0.0367	0.0354	0.0354	0.0370	0.0361
	Φ_5	0.0334	0.0364	0.0361	0.0303	0.0334	0.0329
	Φ_6	0.0295	0.0312	0.0314	0.0306	0.0321	0.0329

Table A2.3: DE Table : BLasso FA, No Model Error

DE : Discrepancy due to Estimation (OLS)

Sample Size	Φ	Model 1			Model 2		
		ML	BLasso (MAP)	BLasso (Posterior Mean)	ML	BLasso (MAP)	BLasso (Posterior Mean)
50	Φ_1	1.0613	1.2279	1.6233	0.9484	0.7985	1.1417
	Φ_2	0.8672	0.7788	1.1074	0.8511	0.6860	0.9393
	Φ_3	0.7355	0.6483	0.8734	0.7578	0.6617	0.9361
	Φ_4	0.6408	0.5684	0.7731	0.6974	0.5515	0.7838
	Φ_5	0.6522	0.7241	1.0397	0.6182	0.6429	0.8894
	Φ_6	0.5730	0.6441	0.8849	0.6049	0.6199	0.8660
100	Φ_1	0.4537	0.4159	0.4675	0.4592	0.4454	0.4861
	Φ_2	0.3557	0.3190	0.3626	0.3898	0.3683	0.4350
	Φ_3	0.3699	0.3472	0.3996	0.3879	0.3579	0.4406
	Φ_4	0.3704	0.3196	0.3772	0.3179	0.2781	0.3370
	Φ_5	0.3076	0.3256	0.4241	0.3187	0.3469	0.4375
	Φ_6	0.2652	0.2889	0.3434	0.2767	0.3009	0.3688
200	Φ_1	0.2171	0.2126	0.2405	0.2388	0.2374	0.2577
	Φ_2	0.2174	0.2119	0.2442	0.1888	0.1860	0.2065
	Φ_3	0.1889	0.1821	0.2011	0.1959	0.1854	0.2060
	Φ_4	0.1729	0.1614	0.1731	0.1674	0.1578	0.1817
	Φ_5	0.1434	0.1663	0.1839	0.1468	0.1561	0.1710
	Φ_6	0.1297	0.1432	0.1553	0.1620	0.1663	0.1958
1000	Φ_1	0.0446	0.0474	0.0486	0.0397	0.0420	0.0411
	Φ_2	0.0397	0.0427	0.0425	0.0400	0.0423	0.0431
	Φ_3	0.0381	0.0375	0.0376	0.0377	0.0388	0.0381
	Φ_4	0.0290	0.0293	0.0288	0.0302	0.0311	0.0306
	Φ_5	0.0313	0.0325	0.0327	0.0295	0.0324	0.0301
	Φ_6	0.0260	0.0287	0.0286	0.0295	0.0309	0.0314

Sample Size	Φ	Model 3			Model 4		
		ML	BLasso (MAP)	BLasso (Posterior Mean)	ML	BLasso (MAP)	BLasso (Posterior Mean)
50	Φ_1	0.9392	0.7857	1.0122	1.1087	0.9816	1.3161
	Φ_2	0.9552	0.7411	1.0205	0.8647	0.7532	1.0173
	Φ_3	0.7192	0.6481	0.8825	0.8667	0.7651	0.9950
	Φ_4	0.7970	0.6837	0.9492	0.7082	0.6395	0.8988
	Φ_5	0.6507	0.7145	1.0191	0.7147	0.8130	1.0705
	Φ_6	0.6094	0.6517	0.9467	0.6517	0.7177	0.9704
100	Φ_1	0.5571	0.5104	0.5745	0.5628	0.5638	0.6268
	Φ_2	0.3582	0.3209	0.3483	0.4256	0.4129	0.4720
	Φ_3	0.3655	0.3487	0.4137	0.4600	0.4448	0.5072
	Φ_4	0.3817	0.3379	0.4143	0.3768	0.3718	0.4501
	Φ_5	0.3137	0.3682	0.4687	0.3458	0.4063	0.4817
	Φ_6	0.3324	0.3692	0.4799	0.3326	0.3762	0.4534
200	Φ_1	0.2435	0.2449	0.2504	0.2945	0.3116	0.3239
	Φ_2	0.2216	0.2155	0.2194	0.2367	0.2431	0.2705
	Φ_3	0.2027	0.2043	0.2211	0.2082	0.2109	0.2214
	Φ_4	0.1835	0.1854	0.1944	0.1780	0.1793	0.1928
	Φ_5	0.1763	0.1923	0.2175	0.1784	0.2066	0.2352
	Φ_6	0.1489	0.1708	0.1910	0.1539	0.1804	0.2064
1000	Φ_1	0.0462	0.0492	0.0480	0.0463	0.0517	0.0530
	Φ_2	0.0407	0.0439	0.0446	0.0388	0.0410	0.0425
	Φ_3	0.0368	0.0378	0.0377	0.0436	0.0442	0.0431
	Φ_4	0.0393	0.0408	0.0396	0.0343	0.0344	0.0352
	Φ_5	0.0351	0.0389	0.0392	0.0345	0.0377	0.0370
	Φ_6	0.0290	0.0297	0.0298	0.0290	0.0310	0.0310

Table A2.4: DE Table : BLasso FA, Model Error Involved

sMSE : Mean Squared Error of Standardized Parameter Estimates

Sample Size	Φ	Model 1			Model 2		
		ML	BLasso (MAP)	BLasso (Posterior Mean)	ML	BLasso (MAP)	BLasso (Posterior Mean)
50	Φ_1	0.0687	0.2143	0.2077	0.0935	0.1069	0.0998
	Φ_2	0.0808	0.1048	0.0971	0.0945	0.1206	0.1112
	Φ_3	0.1898	0.1395	0.1492	0.2453	0.1816	0.2037
	Φ_4	0.2250	0.2543	0.2747	0.2537	0.1735	0.1918
	Φ_5	0.3476	0.4395	0.5057	1.9992	0.5349	0.6094
	Φ_6	0.4925	0.5715	0.6379	0.4783	0.4609	0.5294
100	Φ_1	0.0302	0.0533	0.0502	0.0462	0.0675	0.0650
	Φ_2	0.0400	0.0662	0.0634	0.0481	0.0717	0.0686
	Φ_3	0.0837	0.0854	0.0930	0.1264	0.1057	0.1145
	Φ_4	0.1059	0.0917	0.0959	0.0999	0.0926	0.1014
	Φ_5	0.1515	0.2257	0.2520	0.7544	0.3022	0.3424
	Φ_6	0.2030	0.2524	0.2816	0.2356	0.2849	0.3139
200	Φ_1	0.0151	0.0306	0.0297	0.0188	0.0362	0.0354
	Φ_2	0.0240	0.0402	0.0392	0.0236	0.0400	0.0394
	Φ_3	0.0397	0.0389	0.0410	0.0551	0.0504	0.0528
	Φ_4	0.0445	0.0437	0.0456	0.0562	0.0546	0.0550
	Φ_5	0.0815	0.1083	0.1161	0.1259	0.1377	0.1520
	Φ_6	0.0940	0.1143	0.1259	0.1053	0.1249	0.1367
1000	Φ_1	0.0029	0.0068	0.0067	0.0041	0.0097	0.0096
	Φ_2	0.0040	0.0084	0.0084	0.0044	0.0090	0.0090
	Φ_3	0.0082	0.0083	0.0083	0.0101	0.0099	0.0097
	Φ_4	0.0093	0.0091	0.0090	0.0108	0.0111	0.0109
	Φ_5	0.0150	0.0178	0.0186	0.0269	0.0278	0.0285
	Φ_6	0.0186	0.0210	0.0211	0.0195	0.0217	0.0221

Sample Size	Φ	Model 3			Model 4		
		ML	BLasso (MAP)	BLasso (Posterior Mean)	ML	BLasso (MAP)	BLasso (Posterior Mean)
50	Φ_1	0.2467	0.2283	0.2443	0.0886	0.1354	0.1287
	Φ_2	0.1471	0.2571	0.2530	0.1085	0.1350	0.1300
	Φ_3	0.7190	0.4117	0.4555	0.1976	0.1762	0.1976
	Φ_4	0.3561	0.2450	0.2710	0.2574	0.2644	0.2861
	Φ_5	5.6391	1.0522	1.1649	0.3932	0.4767	0.5374
	Φ_6	1.2695	0.6367	0.7105	0.5231	0.5055	0.5722
100	Φ_1	0.1149	0.1387	0.1442	0.0417	0.0698	0.0678
	Φ_2	0.0650	0.0888	0.0885	0.0543	0.0825	0.0791
	Φ_3	0.3440	0.2183	0.2378	0.1008	0.0984	0.1056
	Φ_4	0.1647	0.1316	0.1428	0.1112	0.1016	0.1101
	Φ_5	12.1045	0.5428	0.5876	0.1579	0.2384	0.2694
	Φ_6	0.3388	0.3178	0.3552	0.2102	0.2481	0.2794
200	Φ_1	0.0476	0.0807	0.0819	0.0219	0.0382	0.0371
	Φ_2	0.0295	0.0461	0.0462	0.0272	0.0436	0.0438
	Φ_3	0.1433	0.1166	0.1263	0.0471	0.0498	0.0508
	Φ_4	0.0731	0.0688	0.0712	0.0574	0.0580	0.0591
	Φ_5	0.5163	0.2868	0.3158	0.0757	0.1150	0.1267
	Φ_6	0.1449	0.1585	0.1739	0.1036	0.1334	0.1466
1000	Φ_1	0.0095	0.0183	0.0183	0.0037	0.0074	0.0074
	Φ_2	0.0060	0.0111	0.0110	0.0052	0.0098	0.0096
	Φ_3	0.0275	0.0266	0.0271	0.0085	0.0090	0.0089
	Φ_4	0.0148	0.0148	0.0147	0.0110	0.0110	0.0110
	Φ_5	0.0604	0.0534	0.0554	0.0155	0.0182	0.0186
	Φ_6	0.0277	0.0286	0.0293	0.0194	0.0211	0.0220

Table A2.5: Standardized MSE Table : BLasso FA, No Model Error

sMSE : Mean Squared Error of Standardized Parameter Estimates

Sample Size	Φ	Model 1			Model 2		
		ML	BLasso (MAP)	BLasso (Posterior Mean)	ML	BLasso (MAP)	BLasso (Posterior Mean)
50	Φ_1	0.0692	0.4203	0.4139	0.0954	0.1188	0.1118
	Φ_2	0.0809	0.2131	0.2091	0.0988	0.1216	0.1129
	Φ_3	0.1834	0.1394	0.1521	0.2790	0.1868	0.2076
	Φ_4	0.2395	0.2381	0.2554	0.2628	0.1715	0.1914
	Φ_5	0.3253	0.4247	0.4880	1.5626	0.5582	0.6255
	Φ_6	0.4491	0.5583	0.6224	1.0555	0.4941	0.5557
100	Φ_1	0.0337	0.0591	0.0555	0.0471	0.0682	0.0647
	Φ_2	0.0407	0.0649	0.0633	0.0492	0.0730	0.0714
	Φ_3	0.0824	0.0781	0.0813	0.1409	0.1188	0.1260
	Φ_4	0.1181	0.0924	0.0975	0.1179	0.0926	0.0998
	Φ_5	0.1695	0.2279	0.2539	0.9917	0.3224	0.3529
	Φ_6	0.1930	0.2239	0.2499	0.2552	0.2637	0.2944
200	Φ_1	0.0150	0.0320	0.0315	0.0275	0.0385	0.0375
	Φ_2	0.0229	0.0417	0.0419	0.0250	0.0404	0.0407
	Φ_3	0.0384	0.0374	0.0388	0.0710	0.0597	0.0601
	Φ_4	0.0550	0.0514	0.0524	0.0618	0.0547	0.0566
	Φ_5	0.0736	0.1063	0.1172	0.1439	0.1569	0.1700
	Φ_6	0.1001	0.1184	0.1290	0.1213	0.1390	0.1515
1000	Φ_1	0.0029	0.0074	0.0074	0.0104	0.0088	0.0087
	Φ_2	0.0039	0.0087	0.0087	0.0074	0.0103	0.0104
	Φ_3	0.0084	0.0081	0.0081	0.0227	0.0213	0.0207
	Φ_4	0.0095	0.0093	0.0092	0.0152	0.0142	0.0138
	Φ_5	0.0153	0.0181	0.0189	0.0339	0.0396	0.0390
	Φ_6	0.0194	0.0218	0.0221	0.0269	0.0299	0.0302

Sample Size	Φ	Model 3			Model 4		
		ML	BLasso (MAP)	BLasso (Posterior Mean)	ML	BLasso (MAP)	BLasso (Posterior Mean)
50	Φ_1	0.2580	0.2137	0.2235	0.0822	0.1174	0.1098
	Φ_2	0.1662	0.1621	0.1579	0.1101	0.1370	0.1331
	Φ_3	0.7628	0.5010	0.5296	0.1985	0.1641	0.1813
	Φ_4	0.3898	0.2713	0.2924	0.2379	0.1828	0.2089
	Φ_5	81.7991	1.2836	1.3810	0.3881	0.4676	0.5272
	Φ_6	1.1118	0.6766	0.7532	0.5869	0.4881	0.5546
100	Φ_1	0.1221	0.1231	0.1243	0.0471	0.0759	0.0725
	Φ_2	0.0723	0.0912	0.0903	0.0528	0.0811	0.0792
	Φ_3	0.3520	0.2907	0.2944	0.1031	0.0925	0.0985
	Φ_4	0.2201	0.1671	0.1752	0.1185	0.1076	0.1167
	Φ_5	1.2674	0.7423	0.7916	0.1778	0.2410	0.2669
	Φ_6	0.4517	0.4232	0.4556	0.2536	0.2731	0.3028
200	Φ_1	0.0763	0.0704	0.0710	0.0212	0.0396	0.0385
	Φ_2	0.0457	0.0539	0.0527	0.0268	0.0462	0.0461
	Φ_3	0.2440	0.2014	0.1973	0.0470	0.0456	0.0470
	Φ_4	0.1298	0.1083	0.1082	0.0592	0.0560	0.0587
	Φ_5	0.6152	0.4617	0.4709	0.0817	0.1238	0.1350
	Φ_6	0.2470	0.2499	0.2603	0.0996	0.1251	0.1360
1000	Φ_1	0.0329	0.0170	0.0169	0.0039	0.0085	0.0084
	Φ_2	0.0163	0.0152	0.0151	0.0045	0.0094	0.0095
	Φ_3	0.1548	0.1420	0.1394	0.0098	0.0097	0.0097
	Φ_4	0.0721	0.0665	0.0656	0.0107	0.0109	0.0108
	Φ_5	0.2853	0.2903	0.2861	0.0157	0.0189	0.0194
	Φ_6	0.1670	0.1639	0.1636	0.0192	0.0209	0.0215

Table A2.6: Standardized MSE Table : BLasso FA, No Model Involved

Variance of Parameter Estimates

Sample Size	Φ	Model 1			Model 2		
		ML	BLasso (MAP)	BLasso (Posterior Mean)	ML	BLasso (MAP)	BLasso (Posterior Mean)
50	Φ_1	0.2259	0.4740	0.5317	0.2797	0.2391	0.2909
	Φ_2	0.2161	0.2052	0.2307	0.2136	0.2010	0.2283
	Φ_3	0.4939	0.3434	0.3628	0.6492	0.4374	0.4764
	Φ_4	0.5157	0.6348	0.6666	0.5825	0.3760	0.3995
	Φ_5	1.1269	0.8829	0.8811	6.1583	1.0459	1.0577
	Φ_6	2.1232	1.3061	1.3130	2.3465	0.9305	0.9247
100	Φ_1	0.1159	0.1225	0.1334	0.1314	0.1335	0.1445
	Φ_2	0.1008	0.1135	0.1205	0.1150	0.1234	0.1311
	Φ_3	0.2155	0.1981	0.2038	0.3020	0.2301	0.2391
	Φ_4	0.2271	0.1911	0.1940	0.2342	0.2028	0.2061
	Φ_5	0.5207	0.4899	0.4712	2.1160	0.6434	0.6434
	Φ_6	0.5663	0.5403	0.5192	0.6424	0.6103	0.5908
200	Φ_1	0.0544	0.0656	0.0697	0.0587	0.0725	0.0755
	Φ_2	0.0562	0.0689	0.0695	0.0578	0.0695	0.0713
	Φ_3	0.1035	0.0971	0.0966	0.1304	0.1185	0.1181
	Φ_4	0.0983	0.0888	0.0880	0.1215	0.1126	0.1105
	Φ_5	0.2461	0.2566	0.2412	0.3721	0.3026	0.2973
	Φ_6	0.2590	0.2676	0.2523	0.2793	0.2883	0.2730
1000	Φ_1	0.0114	0.0152	0.0154	0.0126	0.0178	0.0180
	Φ_2	0.0093	0.0137	0.0139	0.0104	0.0143	0.0144
	Φ_3	0.0210	0.0205	0.0203	0.0237	0.0228	0.0223
	Φ_4	0.0207	0.0206	0.0200	0.0210	0.0209	0.0205
	Φ_5	0.0416	0.0431	0.0414	0.0714	0.0667	0.0653
	Φ_6	0.0467	0.0501	0.0484	0.0496	0.0543	0.0521

Sample Size	Φ	Model 3			Model 4		
		ML	BLasso (MAP)	BLasso (Posterior Mean)	ML	BLasso (MAP)	BLasso (Posterior Mean)
50	Φ_1	0.4715	0.3703	0.4375	0.3145	0.4061	0.4305
	Φ_2	0.3013	0.5002	0.5423	0.2576	0.3574	0.3631
	Φ_3	1.5027	0.7260	0.7908	0.5417	0.5417	0.5682
	Φ_4	0.7387	0.4674	0.4946	0.6411	0.7362	0.7368
	Φ_5	14.7520	1.6842	1.7370	1.5822	1.1042	1.0826
	Φ_6	5.6771	1.2041	1.2078	1.7714	1.1451	1.1226
100	Φ_1	0.2286	0.2279	0.2466	0.1479	0.2058	0.2126
	Φ_2	0.1314	0.1413	0.1542	0.1288	0.1974	0.1956
	Φ_3	0.6820	0.4018	0.4236	0.2551	0.2806	0.2779
	Φ_4	0.3287	0.2549	0.2613	0.2514	0.2742	0.2669
	Φ_5	42.3998	1.0095	1.0064	0.5202	0.5647	0.5440
	Φ_6	0.8856	0.6610	0.6444	0.6415	0.6038	0.5729
200	Φ_1	0.0969	0.1246	0.1313	0.0810	0.1098	0.1094
	Φ_2	0.0635	0.0752	0.0784	0.0741	0.1033	0.1038
	Φ_3	0.2875	0.2233	0.2357	0.1262	0.1476	0.1445
	Φ_4	0.1383	0.1260	0.1257	0.1346	0.1456	0.1413
	Φ_5	1.4418	0.6143	0.6293	0.2540	0.2816	0.2688
	Φ_6	0.3905	0.3479	0.3394	0.2561	0.3044	0.2868
1000	Φ_1	0.0192	0.0266	0.0268	0.0124	0.0175	0.0173
	Φ_2	0.0124	0.0165	0.0166	0.0139	0.0200	0.0197
	Φ_3	0.0538	0.0503	0.0513	0.0229	0.0243	0.0238
	Φ_4	0.0290	0.0286	0.0283	0.0240	0.0251	0.0244
	Φ_5	0.1598	0.1296	0.1313	0.0472	0.0523	0.0503
	Φ_6	0.0672	0.0680	0.0659	0.0499	0.0526	0.0509

Table A2.7: Variance Table : BLasso FA, No Model Error

Variance of Parameter Estimates

Sample Size	Φ	Model 1			Model 2		
		ML	BLasso (MAP)	BLasso (Posterior Mean)	ML	BLasso (MAP)	BLasso (Posterior Mean)
50	Φ_1	0.2713	1.1881	1.2868	0.2584	0.2365	0.2903
	Φ_2	0.2326	0.5026	0.5552	0.2343	0.2244	0.2589
	Φ_3	0.4964	0.3439	0.3670	0.6829	0.4247	0.4705
	Φ_4	0.4838	0.5548	0.5806	0.6345	0.3901	0.4156
	Φ_5	1.1198	0.9173	0.9171	6.5494	1.0971	1.1070
	Φ_6	1.4877	1.3080	1.3073	5.0362	1.0063	0.9978
100	Φ_1	0.1070	0.1180	0.1284	0.1241	0.1383	0.1504
	Φ_2	0.1027	0.1140	0.1230	0.1099	0.1270	0.1377
	Φ_3	0.2248	0.1843	0.1839	0.2847	0.2324	0.2384
	Φ_4	0.2556	0.2095	0.2105	0.2423	0.1961	0.1985
	Φ_5	0.5835	0.4559	0.4415	2.8589	0.6169	0.6142
	Φ_6	0.5490	0.4801	0.4528	0.7196	0.5389	0.5232
200	Φ_1	0.0514	0.0643	0.0681	0.0688	0.0842	0.0890
	Φ_2	0.0558	0.0699	0.0736	0.0514	0.0689	0.0727
	Φ_3	0.1006	0.0959	0.0959	0.1517	0.1296	0.1313
	Φ_4	0.1172	0.1048	0.1026	0.1216	0.1096	0.1099
	Φ_5	0.2173	0.2614	0.2494	0.3804	0.2942	0.2925
	Φ_6	0.2893	0.2702	0.2553	0.3233	0.2771	0.2670
1000	Φ_1	0.0107	0.0145	0.0147	0.0112	0.0172	0.0171
	Φ_2	0.0101	0.0140	0.0140	0.0117	0.0169	0.0170
	Φ_3	0.0218	0.0213	0.0211	0.0257	0.0261	0.0253
	Φ_4	0.0194	0.0193	0.0189	0.0204	0.0210	0.0203
	Φ_5	0.0430	0.0445	0.0434	0.0657	0.0610	0.0608
	Φ_6	0.0461	0.0488	0.0462	0.0551	0.0543	0.0534

Sample Size	Φ	Model 3			Model 4		
		ML	BLasso (MAP)	BLasso (Posterior Mean)	ML	BLasso (MAP)	BLasso (Posterior Mean)
50	Φ_1	0.4626	0.3807	0.4472	0.2888	0.3856	0.4162
	Φ_2	0.2837	0.2590	0.2959	0.2749	0.3785	0.3940
	Φ_3	1.1254	0.7082	0.7703	0.5328	0.4953	0.5058
	Φ_4	0.7080	0.4998	0.5350	0.6107	0.5386	0.5490
	Φ_5	161.4259	1.7314	1.7934	1.2073	1.1214	1.1017
	Φ_6	2.9347	1.0973	1.1048	3.0585	1.1384	1.1131
100	Φ_1	0.1988	0.2274	0.2487	0.1572	0.2202	0.2231
	Φ_2	0.1316	0.1511	0.1590	0.1488	0.2129	0.2129
	Φ_3	0.4662	0.4081	0.4245	0.2629	0.2739	0.2699
	Φ_4	0.3317	0.2663	0.2734	0.2618	0.2823	0.2776
	Φ_5	2.8333	0.9844	1.0322	0.5064	0.5828	0.5543
	Φ_6	0.8424	0.6126	0.6013	0.5782	0.6334	0.6014
200	Φ_1	0.1032	0.1366	0.1448	0.0775	0.1147	0.1121
	Φ_2	0.0670	0.0843	0.0860	0.0719	0.1078	0.1081
	Φ_3	0.2108	0.2006	0.2043	0.1328	0.1466	0.1434
	Φ_4	0.1349	0.1358	0.1331	0.1372	0.1471	0.1437
	Φ_5	0.9259	0.4377	0.4439	0.2316	0.2903	0.2770
	Φ_6	0.3022	0.3105	0.2992	0.2518	0.3051	0.2873
1000	Φ_1	0.0178	0.0303	0.0299	0.0126	0.0180	0.0181
	Φ_2	0.0133	0.0199	0.0199	0.0122	0.0172	0.0173
	Φ_3	0.0366	0.0380	0.0377	0.0258	0.0263	0.0255
	Φ_4	0.0294	0.0298	0.0291	0.0238	0.0244	0.0239
	Φ_5	0.0900	0.0906	0.0904	0.0480	0.0516	0.0501
	Φ_6	0.0621	0.0598	0.0588	0.0500	0.0526	0.0505

Table A2.8: Variance Table : BLasso FA, No Model Involved

OD : Overall Discrepancy (OLS)

Sample Size	Φ	Model 1			Model 2			Model 3		
		ML	BLasso (MAP)	BLasso (Post.Mean)	ML	BLasso (MAP)	BLasso (Post.Mean)	ML	BLasso (MAP)	BLasso (Post.Mean)
150	Φ_1	0.8328	0.8247	0.8018	0.7906	0.9350	0.7858	0.8137	0.8990	0.8092
	Φ_2	0.7107	0.7434	0.6993	0.7024	0.7345	0.7030	0.7710	0.8388	0.8292
	Φ_3	0.6454	0.6914	0.6273	0.6985	0.7447	0.7041	0.6857	0.7576	0.6636
	Φ_4	0.5769	0.6169	0.5836	0.5503	0.6497	0.5605	0.6248	0.6714	0.6021
	Φ_5	0.5304	0.6645	0.5907	0.5253	0.7321	0.6116	0.6025	0.6742	0.6462
	Φ_6	0.5356	0.6613	0.6060	0.5047	0.6981	0.5889	0.5721	0.6541	0.6057
250	Φ_1	0.4951	0.4711	0.4834	0.4820	0.4857	0.4858	0.4689	0.4925	0.4459
	Φ_2	0.4397	0.4485	0.4327	0.4076	0.4143	0.3970	0.4624	0.4582	0.4526
	Φ_3	0.4454	0.4783	0.4613	0.4138	0.4484	0.4061	0.4129	0.4447	0.4023
	Φ_4	0.3437	0.3495	0.3418	0.3698	0.3845	0.3664	0.3660	0.3790	0.3562
	Φ_5	0.3313	0.3824	0.3583	0.3569	0.4069	0.3867	0.3667	0.4393	0.3867
	Φ_6	0.2961	0.3511	0.3185	0.2729	0.3338	0.2980	0.3210	0.3640	0.3313
500	Φ_1	0.2761	0.2793	0.2746	0.2484	0.2558	0.2491	0.2531	0.2819	0.2473
	Φ_2	0.2090	0.2121	0.2041	0.1870	0.1938	0.1848	0.2309	0.2320	0.2245
	Φ_3	0.2008	0.2104	0.2010	0.1936	0.1982	0.1935	0.1971	0.2136	0.1948
	Φ_4	0.1588	0.1753	0.1570	0.1824	0.1972	0.1816	0.1926	0.1990	0.1870
	Φ_5	0.1585	0.1906	0.1709	0.1604	0.1963	0.1658	0.1738	0.1983	0.1795
	Φ_6	0.1468	0.1677	0.1540	0.1432	0.1611	0.1503	0.1546	0.1715	0.1575
1000	Φ_1	0.1243	0.1258	0.1237	0.1403	0.1470	0.1393	0.1274	0.1338	0.1240
	Φ_2	0.1021	0.1052	0.1020	0.1020	0.1058	0.1012	0.1076	0.1108	0.1065
	Φ_3	0.0962	0.1064	0.0977	0.1048	0.1160	0.1038	0.1016	0.1081	0.1015
	Φ_4	0.0856	0.0906	0.0860	0.0890	0.0956	0.0890	0.0920	0.0955	0.0910
	Φ_5	0.0852	0.0953	0.0899	0.0871	0.1007	0.0902	0.0839	0.1022	0.0862
	Φ_6	0.0787	0.0828	0.0821	0.0739	0.0812	0.0773	0.0799	0.0906	0.0811

Table A2.9: OD Table : BLasso SEM, No Model Error

OD : Overall Discrepancy (OLS)

Sample Size	Φ	Model 1			Model 2			Model 3		
		ML	BLasso (MAP)	BLasso (Post.Mean)	ML	BLasso (MAP)	BLasso (Post.Mean)	ML	BLasso (MAP)	BLasso (Post.Mean)
150	Φ_1	0.8103	0.8473	0.7615	0.8621	1.0277	0.8660	0.8910	0.9469	0.8751
	Φ_2	0.8108	0.8733	0.8031	0.7912	0.8956	0.8163	0.8139	0.8793	0.7760
	Φ_3	0.7327	0.8759	0.7794	0.8256	1.0177	0.8928	0.7631	0.8917	0.7564
	Φ_4	0.6943	0.8425	0.7281	0.6690	0.9133	0.7544	0.7096	0.8325	0.7634
	Φ_5	0.6107	0.8081	0.7207	0.6509	0.9617	0.8083	0.6447	0.7654	0.7008
	Φ_6	0.6091	0.8664	0.7337	0.6250	0.9174	0.7657	0.5294	0.6753	0.5930
250	Φ_1	0.5676	0.5940	0.5367	0.5907	0.6430	0.5544	0.5601	0.6040	0.5470
	Φ_2	0.5301	0.5800	0.5204	0.5047	0.5214	0.4995	0.4331	0.4425	0.4151
	Φ_3	0.5470	0.6730	0.5700	0.4196	0.5003	0.4357	0.3918	0.4691	0.3941
	Φ_4	0.4752	0.5748	0.4999	0.4778	0.5613	0.5021	0.3637	0.3966	0.3586
	Φ_5	0.3901	0.4972	0.4649	0.3932	0.5354	0.4595	0.3719	0.4993	0.4078
	Φ_6	0.4092	0.5475	0.4837	0.4131	0.5463	0.4815	0.3662	0.4436	0.3924
500	Φ_1	0.2727	0.2988	0.2616	0.2804	0.3006	0.2599	0.2616	0.2947	0.2552
	Φ_2	0.2676	0.2857	0.2625	0.2932	0.3156	0.2862	0.2539	0.2676	0.2476
	Φ_3	0.2556	0.3101	0.2693	0.2653	0.3164	0.2765	0.2163	0.2650	0.2171
	Φ_4	0.2640	0.3184	0.2822	0.2720	0.3238	0.2857	0.2239	0.2439	0.2254
	Φ_5	0.2314	0.3135	0.2674	0.2430	0.3278	0.2766	0.1938	0.2330	0.2061
	Φ_6	0.2390	0.3228	0.2827	0.2385	0.3102	0.2724	0.1746	0.2016	0.1865
1000	Φ_1	0.1777	0.1914	0.1710	0.1660	0.1784	0.1590	0.1459	0.1531	0.1448
	Φ_2	0.1817	0.1979	0.1808	0.1766	0.1850	0.1740	0.1189	0.1290	0.1155
	Φ_3	0.1793	0.2108	0.1911	0.1619	0.1970	0.1717	0.1183	0.1452	0.1201
	Φ_4	0.1809	0.2132	0.1957	0.1838	0.2130	0.1953	0.1159	0.1268	0.1180
	Φ_5	0.1451	0.1835	0.1674	0.1585	0.2015	0.1793	0.1014	0.1183	0.1082
	Φ_6	0.1561	0.1945	0.1796	0.1754	0.2119	0.1952	0.0918	0.1057	0.0978

Table A2.10: OD Table : BLasso SEM, Model Error Involved

DE : Discrepancy due to Estimation (OLS)

Sample Size	Φ	Model 1			Model 2			Model 3		
		ML	BLasso (MAP)	BLasso (Post.Mean)	ML	BLasso (MAP)	BLasso (Post.Mean)	ML	BLasso (MAP)	BLasso (Post.Mean)
150	Φ_1	0.8328	0.8280	0.8000	0.7906	0.9197	0.7831	0.8137	0.8884	0.8077
	Φ_2	0.7107	0.7409	0.6973	0.7024	0.7281	0.7008	0.7710	0.8338	0.8277
	Φ_3	0.6453	0.6867	0.6248	0.6985	0.7397	0.7010	0.6857	0.7541	0.6615
	Φ_4	0.5769	0.6142	0.5813	0.5503	0.6398	0.5574	0.6248	0.6576	0.5995
	Φ_5	0.5304	0.6584	0.5861	0.5253	0.7163	0.6058	0.5905	0.6425	0.6290
	Φ_6	0.5356	0.6538	0.6012	0.5047	0.6825	0.5834	0.5721	0.6408	0.6019
250	Φ_1	0.4951	0.4720	0.4832	0.4820	0.4791	0.4848	0.4689	0.4856	0.4450
	Φ_2	0.4397	0.4469	0.4318	0.4076	0.4110	0.3962	0.4624	0.4564	0.4520
	Φ_3	0.4454	0.4752	0.4595	0.4138	0.4457	0.4047	0.4129	0.4425	0.4013
	Φ_4	0.3437	0.3482	0.3406	0.3698	0.3796	0.3651	0.3660	0.3710	0.3546
	Φ_5	0.3313	0.3779	0.3554	0.3533	0.3940	0.3799	0.3630	0.4215	0.3799
	Φ_6	0.2961	0.3462	0.3156	0.2729	0.3251	0.2952	0.3210	0.3551	0.3285
500	Φ_1	0.2761	0.2799	0.2741	0.2484	0.2526	0.2484	0.2531	0.2769	0.2467
	Φ_2	0.2090	0.2109	0.2035	0.1870	0.1923	0.1844	0.2309	0.2301	0.2238
	Φ_3	0.2008	0.2086	0.2001	0.1936	0.1965	0.1927	0.1971	0.2125	0.1943
	Φ_4	0.1588	0.1741	0.1558	0.1824	0.1938	0.1807	0.1926	0.1947	0.1861
	Φ_5	0.1585	0.1875	0.1687	0.1604	0.1900	0.1639	0.1738	0.1902	0.1778
	Φ_6	0.1467	0.1648	0.1522	0.1432	0.1566	0.1486	0.1546	0.1658	0.1558
1000	Φ_1	0.1243	0.1266	0.1233	0.1403	0.1451	0.1390	0.1274	0.1310	0.1235
	Φ_2	0.1021	0.1046	0.1017	0.1020	0.1046	0.1008	0.1076	0.1099	0.1062
	Φ_3	0.0962	0.1050	0.0969	0.1048	0.1149	0.1030	0.1016	0.1076	0.1011
	Φ_4	0.0856	0.0901	0.0854	0.0890	0.0936	0.0883	0.0920	0.0929	0.0904
	Φ_5	0.0852	0.0935	0.0886	0.0871	0.0974	0.0889	0.0839	0.0970	0.0851
	Φ_6	0.0787	0.0811	0.0811	0.0739	0.0784	0.0762	0.0799	0.0867	0.0798

Table A2.11: DE Table : BLasso SEM, No Model Error

DE : Discrepancy due to Estimation (OLS)

Sample Size	Φ	Model 1			Model 2			Model 3		
		ML	BLasso (MAP)	BLasso (Post.Mean)	ML	BLasso (MAP)	BLasso (Post.Mean)	ML	BLasso (MAP)	BLasso (Post.Mean)
150	Φ_1	0.7382	0.7272	0.7071	0.7990	0.8760	0.7959	0.8738	0.9156	0.8656
	Φ_2	0.7328	0.7249	0.7230	0.7283	0.7376	0.7364	0.7840	0.8348	0.7507
	Φ_3	0.6663	0.7171	0.6750	0.7769	0.8474	0.7889	0.7511	0.8523	0.7368
	Φ_4	0.5757	0.6062	0.5597	0.5587	0.6835	0.5805	0.6958	0.8031	0.7424
	Φ_5	0.5540	0.6426	0.5956	0.5861	0.7861	0.6637	0.6367	0.7196	0.6751
	Φ_6	0.4986	0.6330	0.5458	0.5147	0.6886	0.5766	0.5207	0.6548	0.5703
250	Φ_1	0.5033	0.5029	0.4897	0.5128	0.5374	0.4945	0.5444	0.5831	0.5398
	Φ_2	0.4494	0.4556	0.4454	0.4376	0.4193	0.4421	0.4162	0.4185	0.4040
	Φ_3	0.4658	0.5265	0.4623	0.3547	0.3732	0.3487	0.3731	0.4324	0.3713
	Φ_4	0.3709	0.3897	0.3633	0.3725	0.3921	0.3733	0.3482	0.3708	0.3393
	Φ_5	0.3366	0.3704	0.3658	0.3297	0.4015	0.3516	0.3578	0.4583	0.3814
	Φ_6	0.3311	0.3789	0.3525	0.3277	0.3796	0.3496	0.3543	0.4232	0.3711
500	Φ_1	0.2197	0.2340	0.2188	0.2141	0.2218	0.2056	0.2483	0.2783	0.2463
	Φ_2	0.1923	0.1919	0.1924	0.2133	0.2147	0.2123	0.2371	0.2464	0.2338
	Φ_3	0.1832	0.1997	0.1810	0.2041	0.2142	0.2008	0.1977	0.2362	0.1951
	Φ_4	0.1675	0.1726	0.1654	0.1748	0.1864	0.1727	0.2100	0.2236	0.2085
	Φ_5	0.1587	0.1916	0.1649	0.1721	0.2119	0.1776	0.1808	0.2038	0.1861
	Φ_6	0.1427	0.1704	0.1532	0.1487	0.1703	0.1545	0.1652	0.1859	0.1712
1000	Φ_1	0.1199	0.1260	0.1190	0.1082	0.1129	0.1072	0.1345	0.1399	0.1360
	Φ_2	0.0975	0.1000	0.0972	0.0955	0.0962	0.0958	0.1007	0.1079	0.0984
	Φ_3	0.1084	0.1168	0.1083	0.0941	0.1054	0.0937	0.1014	0.1213	0.1001
	Φ_4	0.0879	0.0900	0.0878	0.0831	0.0890	0.0829	0.1015	0.1079	0.1009
	Φ_5	0.0792	0.0867	0.0818	0.0895	0.1047	0.0922	0.0872	0.0944	0.0892
	Φ_6	0.0708	0.0725	0.0730	0.0800	0.0873	0.0817	0.0817	0.0911	0.0835

Table A2.12: DE Table : BLasso SEM, Model Error Involved

sMSE : Mean Squared Error of Standardized Parameter Estimates

Sample Size	Φ	Model 1			Model 2			Model 3		
		ML	BLasso (MAP)	BLasso (Post.Mean)	ML	BLasso (MAP)	BLasso (Post.Mean)	ML	BLasso (MAP)	BLasso (Post.Mean)
150	Φ_1	0.1005	0.0882	0.0851	0.1251	0.1219	0.1160	0.2104	0.1607	0.1638
	Φ_2	0.1116	0.1001	0.0973	0.1378	0.1249	0.1217	0.1934	0.2446	0.2460
	Φ_3	0.1637	0.1326	0.1300	0.2157	0.1710	0.1711	0.3858	0.2375	0.2442
	Φ_4	0.1819	0.1584	0.1570	0.2210	0.1872	0.1830	0.2695	0.2152	0.2172
	Φ_5	0.2504	0.2924	0.2968	0.4498	0.3587	0.3621	1.1297	0.4323	0.4516
	Φ_6	0.3051	0.3542	0.3572	0.3946	0.3990	0.3918	0.4809	0.4095	0.4206
250	Φ_1	0.0543	0.0507	0.0492	0.0746	0.0709	0.0706	0.1204	0.0913	0.0928
	Φ_2	0.0671	0.0616	0.0607	0.0848	0.0718	0.0714	0.1149	0.0952	0.0947
	Φ_3	0.0938	0.0888	0.0874	0.1251	0.1089	0.1103	0.2041	0.1457	0.1504
	Φ_4	0.1027	0.0984	0.0985	0.1377	0.1170	0.1164	0.1706	0.1328	0.1351
	Φ_5	0.1513	0.1729	0.1763	0.1992	0.1987	0.2066	0.4668	0.2808	0.2912
	Φ_6	0.1699	0.1953	0.1984	0.1924	0.2100	0.2147	0.2605	0.2470	0.2517
500	Φ_1	0.0278	0.0276	0.0270	0.0342	0.0320	0.0315	0.0548	0.0459	0.0469
	Φ_2	0.0318	0.0308	0.0306	0.0380	0.0358	0.0352	0.0543	0.0483	0.0489
	Φ_3	0.0469	0.0453	0.0445	0.0579	0.0525	0.0528	0.1039	0.0792	0.0811
	Φ_4	0.0524	0.0506	0.0495	0.0615	0.0586	0.0583	0.0845	0.0720	0.0731
	Φ_5	0.0745	0.0859	0.0866	0.0965	0.1072	0.1088	0.1785	0.1346	0.1394
	Φ_6	0.0858	0.0945	0.0957	0.0987	0.1070	0.1088	0.1263	0.1233	0.1258
1000	Φ_1	0.0133	0.0137	0.0134	0.0174	0.0175	0.0172	0.0288	0.0248	0.0249
	Φ_2	0.0159	0.0165	0.0161	0.0197	0.0195	0.0191	0.0267	0.0251	0.0251
	Φ_3	0.0229	0.0233	0.0226	0.0277	0.0271	0.0264	0.0446	0.0390	0.0398
	Φ_4	0.0265	0.0268	0.0265	0.0310	0.0307	0.0305	0.0369	0.0358	0.0358
	Φ_5	0.0369	0.0425	0.0420	0.0509	0.0520	0.0521	0.0842	0.0703	0.0716
	Φ_6	0.0417	0.0488	0.0494	0.0468	0.0523	0.0529	0.0606	0.0634	0.0635

Table A2.13: Standardized MSE Table : BLasso SEM, No Model Error

sMSE : Mean Squared Error of Standardized Parameter Estimates

Sample Size	Φ	Model 1			Model 2			Model 3		
		ML	BLasso (MAP)	BLasso (Post.Mean)	ML	BLasso (MAP)	BLasso (Post.Mean)	ML	BLasso (MAP)	BLasso (Post.Mean)
150	Φ_1	0.0952	0.0818	0.0780	0.1299	0.1357	0.1320	0.2552	0.2167	0.2210
	Φ_2	0.1170	0.1038	0.1011	0.1481	0.1485	0.1457	0.2329	0.2211	0.2198
	Φ_3	0.1571	0.1347	0.1337	0.2193	0.1975	0.1919	0.4938	0.4482	0.4501
	Φ_4	0.1821	0.1529	0.1507	0.2320	0.2210	0.2162	0.4309	0.4544	0.4566
	Φ_5	0.2497	0.2898	0.2988	1.7742	0.3998	0.4002	0.9786	0.7170	0.7241
	Φ_6	0.2945	0.3339	0.3352	0.3685	0.4407	0.4361	0.6068	0.6006	0.6090
250	Φ_1	0.0565	0.0544	0.0529	0.0735	0.0765	0.0747	0.1925	0.1688	0.1687
	Φ_2	0.0675	0.0642	0.0627	0.0892	0.0940	0.0935	0.1491	0.1459	0.1471
	Φ_3	0.1028	0.0919	0.0888	0.1373	0.1303	0.1307	0.3658	0.3549	0.3552
	Φ_4	0.1101	0.1007	0.0996	0.1464	0.1459	0.1448	0.2566	0.2520	0.2520
	Φ_5	0.1458	0.1839	0.1891	0.2317	0.2526	0.2581	0.9688	0.6255	0.6238
	Φ_6	0.1661	0.2013	0.2056	0.2358	0.2752	0.2779	0.4460	0.4561	0.4627
500	Φ_1	0.0268	0.0275	0.0264	0.0394	0.0419	0.0417	0.1199	0.1058	0.1052
	Φ_2	0.0324	0.0311	0.0309	0.0498	0.0536	0.0525	0.0935	0.0950	0.0951
	Φ_3	0.0449	0.0435	0.0426	0.0713	0.0745	0.0746	0.2675	0.2674	0.2661
	Φ_4	0.0522	0.0503	0.0499	0.0803	0.0847	0.0838	0.1925	0.1978	0.1967
	Φ_5	0.0722	0.0857	0.0852	0.1329	0.1451	0.1462	0.4915	0.4910	0.4936
	Φ_6	0.0832	0.0996	0.1003	0.1237	0.1445	0.1453	0.3078	0.3185	0.3202
1000	Φ_1	0.0133	0.0138	0.0132	0.0222	0.0240	0.0238	0.0912	0.0819	0.0817
	Φ_2	0.0161	0.0165	0.0162	0.0319	0.0347	0.0345	0.0674	0.0683	0.0679
	Φ_3	0.0228	0.0231	0.0222	0.0440	0.0477	0.0477	0.2189	0.2220	0.2204
	Φ_4	0.0256	0.0264	0.0257	0.0567	0.0612	0.0601	0.1385	0.1432	0.1423
	Φ_5	0.0350	0.0413	0.0409	0.0774	0.0898	0.0906	0.3860	0.3974	0.3964
	Φ_6	0.0410	0.0470	0.0478	0.0794	0.0896	0.0890	0.2287	0.2389	0.2372

Table A2.14: Standardized MSE Table : BLasso SEM, No Model Involved

Variance of Parameter Estimates

Sample Size	Φ	Model 1			Model 2			Model 3		
		ML	BLasso (MAP)	BLasso (Post.Mean)	ML	BLasso (MAP)	BLasso (Post.Mean)	ML	BLasso (MAP)	BLasso (Post.Mean)
150	Φ_1	0.2001	0.1649	0.1696	0.2069	0.1911	0.1980	0.2927	0.2356	0.2458
	Φ_2	0.1997	0.1733	0.1727	0.2224	0.2051	0.2071	0.2761	0.3887	0.3945
	Φ_3	0.3195	0.2666	0.2589	0.3693	0.3299	0.3318	0.5350	0.3584	0.3649
	Φ_4	0.3279	0.2998	0.2867	0.3614	0.3456	0.3363	0.4241	0.3439	0.3362
	Φ_5	0.5152	0.5971	0.5723	0.8556	0.7230	0.7030	1.6014	0.7513	0.7375
	Φ_6	0.6873	0.7506	0.7183	0.8029	0.8230	0.7875	0.8703	0.7412	0.7175
250	Φ_1	0.1156	0.1005	0.1009	0.1304	0.1197	0.1218	0.1690	0.1325	0.1355
	Φ_2	0.1187	0.1061	0.1053	0.1348	0.1136	0.1130	0.1652	0.1370	0.1372
	Φ_3	0.1672	0.1683	0.1658	0.2206	0.1916	0.1919	0.3003	0.2274	0.2280
	Φ_4	0.1910	0.1811	0.1765	0.2277	0.2008	0.1947	0.2515	0.2092	0.2063
	Φ_5	0.3224	0.3664	0.3537	0.3896	0.3882	0.3781	0.7652	0.4714	0.4647
	Φ_6	0.3524	0.3996	0.3807	0.3803	0.4246	0.4060	0.4935	0.4694	0.4523
500	Φ_1	0.0570	0.0551	0.0553	0.0644	0.0600	0.0602	0.0788	0.0681	0.0698
	Φ_2	0.0577	0.0543	0.0535	0.0635	0.0593	0.0591	0.0765	0.0685	0.0686
	Φ_3	0.0867	0.0860	0.0835	0.1031	0.0936	0.0932	0.1467	0.1189	0.1197
	Φ_4	0.0919	0.0915	0.0884	0.1055	0.0977	0.0956	0.1279	0.1126	0.1121
	Φ_5	0.1479	0.1718	0.1660	0.1971	0.2015	0.1975	0.3033	0.2249	0.2222
	Φ_6	0.1675	0.1826	0.1739	0.1932	0.2157	0.2100	0.2136	0.2223	0.2155
1000	Φ_1	0.0267	0.0270	0.0265	0.0333	0.0317	0.0317	0.0410	0.0364	0.0365
	Φ_2	0.0281	0.0283	0.0278	0.0330	0.0322	0.0318	0.0379	0.0360	0.0359
	Φ_3	0.0429	0.0440	0.0433	0.0471	0.0457	0.0446	0.0636	0.0578	0.0582
	Φ_4	0.0476	0.0494	0.0479	0.0517	0.0512	0.0503	0.0545	0.0543	0.0533
	Φ_5	0.0720	0.0815	0.0792	0.0923	0.0959	0.0933	0.1377	0.1161	0.1154
	Φ_6	0.0829	0.0979	0.0954	0.0869	0.0998	0.0973	0.1031	0.1101	0.1073

Table A2.15: Variance Table : BLasso SEM, No Model Error

Variance of Parameter Estimates

Sample Size	Φ	Model 1			Model 2			Model 3		
		ML	BLasso (MAP)	BLasso (Post.Mean)	ML	BLasso (MAP)	BLasso (Post.Mean)	ML	BLasso (MAP)	BLasso (Post.Mean)
150	Φ_1	0.1912	0.1592	0.1608	0.2162	0.1911	0.1943	0.2736	0.2139	0.2194
	Φ_2	0.1939	0.1695	0.1693	0.2297	0.2074	0.2084	0.2593	0.2114	0.2102
	Φ_3	0.2909	0.2732	0.2672	0.3645	0.3132	0.3086	0.4832	0.3479	0.3403
	Φ_4	0.3296	0.2890	0.2788	0.3602	0.3473	0.3350	0.4317	0.5187	0.5110
	Φ_5	0.5367	0.5764	0.5531	2.9301	0.6807	0.6550	1.1988	0.6460	0.6157
	Φ_6	0.6565	0.7011	0.6641	0.6738	0.7936	0.7537	0.7366	0.7273	0.6883
250	Φ_1	0.1106	0.0991	0.0986	0.1278	0.1173	0.1175	0.1692	0.1461	0.1466
	Φ_2	0.1213	0.1104	0.1096	0.1286	0.1181	0.1177	0.1485	0.1288	0.1284
	Φ_3	0.1829	0.1656	0.1617	0.2075	0.1772	0.1730	0.2629	0.2192	0.2147
	Φ_4	0.1959	0.1820	0.1752	0.2099	0.1966	0.1881	0.2291	0.2002	0.1909
	Φ_5	0.2979	0.3717	0.3586	0.3589	0.3620	0.3450	1.1163	0.4187	0.3991
	Φ_6	0.3252	0.4248	0.4072	0.3540	0.4532	0.4328	0.4530	0.4473	0.4249
500	Φ_1	0.0508	0.0498	0.0495	0.0579	0.0545	0.0543	0.0773	0.0711	0.0712
	Φ_2	0.0566	0.0538	0.0532	0.0600	0.0592	0.0578	0.0756	0.0701	0.0691
	Φ_3	0.0809	0.0784	0.0763	0.1007	0.0939	0.0920	0.1278	0.1139	0.1111
	Φ_4	0.0929	0.0900	0.0874	0.0982	0.0969	0.0931	0.1161	0.1097	0.1059
	Φ_5	0.1432	0.1663	0.1575	0.1991	0.1870	0.1797	0.2445	0.2141	0.2067
	Φ_6	0.1637	0.1985	0.1902	0.1836	0.2156	0.2065	0.2056	0.2070	0.1974
1000	Φ_1	0.0270	0.0270	0.0269	0.0282	0.0279	0.0275	0.0369	0.0375	0.0367
	Φ_2	0.0282	0.0281	0.0277	0.0316	0.0318	0.0314	0.0377	0.0364	0.0357
	Φ_3	0.0430	0.0438	0.0426	0.0487	0.0473	0.0463	0.0597	0.0568	0.0554
	Φ_4	0.0437	0.0457	0.0445	0.0474	0.0502	0.0484	0.0561	0.0546	0.0529
	Φ_5	0.0730	0.0842	0.0811	0.0887	0.0973	0.0937	0.1092	0.1063	0.1031
	Φ_6	0.0798	0.0948	0.0920	0.0939	0.1025	0.0992	0.0954	0.1037	0.0990

Table A2.16: Variance Table : BLasso SEM, No Model Involved

Appendix B : Standardization of SEM

General SEM programs produce two types of estimates - unstandardized and standardized. The former is the estimates which include several constraints to enable SEM estimation. And the latter is the rescaled version of unstandardized estimates. Note that this standardization is different from that in linear regression analysis, which indicates just centering and rescaling the variables before the analysis is conducted. When we study on the estimation of SEM, this two kinds of estimates can be the issue. In this appendix, the formula representing standardization of SEM will be presented. Here, operator $D(\cdot)$ indicates $Diag(\cdot)$, the matrix-valued function which produce a diagonal matrix containing only the diagonal elements in the input matrix. And a supindex s represent that the component is a standardized estimate.

B1. Factor Analysis Model

The original factor analysis model with its estimate matrix is as follows.

$$\hat{\Sigma} = \hat{\Lambda}\hat{\Phi}\hat{\Lambda}^T + \hat{\Psi}_\epsilon$$

Standardization of the model produces the scaled implied covariance matrix, which can be called as implied correlation matrix.

$$\begin{aligned}
\hat{\Sigma}^s &= D(\hat{\Sigma})^{-1/2}\ \hat{\Sigma}\ D(\hat{\Sigma})^{-1/2} \\
&= D(\hat{\Sigma})^{-1/2}\ \hat{\Lambda}\ \hat{\Phi}\hat{\Lambda}^T\ D(\hat{\Sigma})^{-1/2} + D(\hat{\Sigma})^{-1/2}\ \hat{\Psi}_\epsilon\ D(\hat{\Sigma})^{-1/2} \\
&= D(\hat{\Sigma})^{-1/2}\ \hat{\Lambda}\ D(\hat{\Phi})^{1/2}\ D(\hat{\Phi})^{-1/2}\ \hat{\Phi}\ D(\hat{\Phi})^{-1/2}\ D(\hat{\Phi})^{1/2}\hat{\Lambda}^T\ D(\hat{\Sigma})^{-1/2} \\
&\quad + D(\hat{\Sigma})^{-1/2}\ \hat{\Psi}_\epsilon\ D(\hat{\Sigma})^{-1/2} \\
&= \hat{\Lambda}^s\hat{\Phi}\hat{\Lambda}^{sT} + \hat{\Psi}^s_\epsilon
\end{aligned}$$

Hence, standardization of matrices can be done with following formulas.

$$\hat{\Lambda}^s = D(\hat{\Sigma})^{-1/2}\ \hat{\Lambda}\ D(\hat{\Phi})^{1/2}$$
$$\hat{\Psi}^s_\epsilon = D(\hat{\Sigma})^{-1/2}\ \hat{\Psi}_\epsilon\ D(\hat{\Sigma})^{-1/2}$$
$$\hat{\Phi}^s = D(\hat{\Phi})^{-1/2}\ \hat{\Phi}\ D(\hat{\Phi})^{-1/2}$$

B2. Structural Equation Model

Standardization of general structural equation modeling can also be carried out under the same logic. Here only the result will be posed. $\hat{C}$ is the covariance matrix among the latent endogenous variables(η), computed as follows.

$$\hat{C}\ = COV(\eta) = (I-\hat{B})^{-1}(\hat{\Gamma}\hat{\Phi}\hat{\Gamma}' + \hat{\Psi}_\zeta)((I-\hat{B})^{-1})'$$

Measurement Model

$$\hat{\Lambda}^s_y = D(\hat{\Sigma}_{yy})^{-1/2}\ \hat{\Lambda}_y\ D(\hat{C})^{1/2}$$
$$\hat{\Lambda}^s_x = D(\hat{\Sigma}_{xx})^{-1/2}\ \hat{\Lambda}_x\ D(\hat{\Phi})^{1/2}$$
$$\hat{\Theta}^s_y = D(\hat{\Sigma}_{yy})^{-1/2}\ \hat{\Theta}_y\ D(\hat{\Sigma}_{yy})^{-1/2}$$
$$\hat{\Theta}^s_x = D(\hat{\Sigma}_{xx})^{-1/2}\ \hat{\Theta}_x\ D(\hat{\Sigma}_{xx})^{-1/2}$$
$$\hat{\Phi}^s = D(\hat{\Phi})^{-1/2}\ \hat{\Phi}\ D(\hat{\Phi})^{-1/2}$$

Structural Model

$$\hat{B}^s = D(\hat{C})^{-1/2}\ \hat{B}\ D(\hat{C})^{1/2}$$
$$\hat{\Gamma}^s = D(\hat{C})^{-1/2}\ \hat{\Gamma}\ D(\hat{\Phi})^{1/2}$$
$$\hat{\Psi}^s_\zeta = D(\hat{C})^{-1/2}\ \hat{\Psi}_\zeta\ D(\hat{C})^{-1/2}$$

Appendix C : Some Derivations

C1. Derivation of Latent Variable Covariance Matrix Σ_ω

Formula (1.1.10) for covariance matrix of latent variables Σ_ω can be derived as follows. We start from the model (1.1.6) $\sim$ (1.1.7).

Since we know $\Sigma_{\xi\xi} = \Phi$ for the variance matrix of ξ, we only have to obtain the variance matrix of η and the covariance between η and ξ. First, if we transport the first term in RHS to LHS,

$$(I - B)\eta_i = \Gamma\xi_i + \zeta_i \ \sim \ N_{q_1}(0, \Gamma\Phi\Gamma^T + \Psi_\zeta)$$

Hence,

$$\begin{aligned} &\text{var}((I - B)\eta_i) = \Gamma\Phi\Gamma^T + \Psi_\zeta \\ &\Rightarrow \ \Sigma_{\eta\eta} = \text{var}(\eta_i) = (I - B)^{-1}(\Gamma\Phi\Gamma^T + \Psi_\zeta)((I - B)^{-1})^T \end{aligned}$$

And,

$$\begin{aligned} \text{cov}((I - B)\eta_i,\ {\xi_i}^T) &= \text{cov}(\Gamma\xi_i + \zeta_i,\ {\xi_i}^T) \\ &= \Gamma\ \text{cov}(\xi_i,\ {\xi_i}^T) + \text{cov}(\zeta_i,\ {\xi_i}^T) = \Gamma\Phi \\ \Rightarrow \ \Sigma_{\eta\xi} &= \text{cov}(\eta_i,\ {\xi_i}^T) = (I - B)^{-1}\Gamma\Phi \\ \Rightarrow \ \Sigma_{\xi\eta} &= \Sigma_{\eta\xi}^T \end{aligned}$$

C2. Derivations for Some Posterior Distributions of BLasso SEM

In Chapter 3, prior and posterior distributions for many cases, including basic or regularized, and FA or SEM, are represented. Most of derivations for posterior distributions are described in the previous researches. For example, results for basic Bayesian FA and SEM are derived gently in Song and Lee(2012a).

Results for BLasso SEM are also described in Guo et al(2012). However, their result regularized only the structural model, and the measurement part remained intact. Also, the derivations for regularizing parts of posterior distributions are not described in their paper and supplements, with only the results suggested. Hence, in this section, we shall give details on those derivations.

The prior distribution for BLasso FA and SEM can be seen in Section 3.2. For the more broader SEM model, we need to derive posterior distributions for 1) Latent Score(ω_i), 2) Measurement Model part($\Psi_\epsilon, \Lambda, \tau_{\Lambda_j}, \kappa^2_{\Lambda_j}$), and 3) Structural Model part($\Psi_\zeta, \Lambda_\eta, \tau_{\Lambda_{\eta k}}, \kappa^2_{\Lambda_{\eta k}}, \Phi$). Among those terms, the results of ω_i, Ψ_ϵ, Λ, Ψ_ζ, Λ_η, and Φ have only slight difference with the basic Bayesian SEM case, whose derivations are in Song and Lee(2012a). What makes distinction is the remaining parameters, $\tau_{\Lambda_j}, \kappa^2_{\Lambda_j}, \tau_{\Lambda_{\eta k}}$, and $\kappa^2_{\Lambda_{\eta k}}$. Studies on these terms are important since they are introduced with the purpose of regularization in BLasso SEM.

i) Posterior Distribution of τ_{Λ_j}

$$
\begin{aligned}
&P(\tau_{\Lambda_j}|Y,\theta) \\
&\propto P(\Lambda_j|\psi_{\epsilon j},\tau_{\Lambda j})P(\tau_{\Lambda j}|\kappa^2_{\Lambda_j}) \\
&\propto |\psi_{\epsilon j}H_{0yj}|^{-1/2}\exp(-\frac{1}{2}\psi^{-1}_{\epsilon j}(\Lambda_j-\Lambda_{0j})^T H^{-1}_{0yj}(\Lambda_j-\Lambda_{0j}))\times\prod_k^{q_{(j)}}\frac{\kappa^2_{\Lambda_j}}{2}\exp(-\frac{\kappa^2_{\Lambda_j}\tau^2_{\Lambda_{jk}}}{2})
\end{aligned}
$$

$$
\begin{aligned}
&\propto (\prod_{k}^{q_{(j)}} \tau^2_{\Lambda_{jk}})^{-1/2} \text{exp}(\sum_{k}^{q_{(j)}} [-\frac{1}{2}\psi^{-1}_{\epsilon j}(\tau^2_{\Lambda_{jk}})^{-1}(\Lambda_{jk}-\Lambda_{0jk})^2]) \times \prod_{k}^{q_{(j)}} \text{exp}(-\frac{\kappa^2_{\Lambda_j}\tau^2_{\Lambda_{jk}}}{2}) \\
&= \prod_{k}^{q_{(j)}} \left[(\tau^2_{\Lambda_{jk}})^{-1/2} \text{exp}(-\frac{1}{2}\psi^{-1}_{\epsilon j}(\tau^2_{\Lambda_{jk}})^{-1}(\Lambda_{jk}-\Lambda_{0jk})^2) \times \text{exp}(-\frac{\kappa^2_{\Lambda_j}\tau^2_{\Lambda_{jk}}}{2}) \right] \\
&\propto \prod_{k}^{q_{(j)}} P(\tau^2_{\Lambda_{jk}} | \Lambda_{jk}, \psi_{\epsilon j}, \kappa^2_{\Lambda_j})
\end{aligned}
$$

The last line above shows that the posterior distribution of vector τ_{Λ_j} can be decomposed into those of each $\tau^2_{\Lambda_{jk}}$, and can be dealt with independently.

$$
\begin{aligned}
&P(\tau^2_{\Lambda_{jk}} | \Lambda_{jk}, \psi_{\epsilon j}, \kappa^2_{\Lambda_j}) \\
&\propto (\tau^2_{\Lambda_{jk}})^{-1/2} \text{exp}(-\frac{1}{2}\psi^{-1}_{\epsilon j}(\tau^2_{\Lambda_{jk}})^{-1}(\Lambda_{jk}-\Lambda_{0jk})^2) \times \text{exp}(-\frac{\kappa^2_{\Lambda_j}\tau^2_{\Lambda_{jk}}}{2}) \\
&\propto (\tau^2_{\Lambda_{jk}})^{-1/2} \text{exp}(-\frac{1}{2}\psi^{-1}_{\epsilon j}(\tau_{\Lambda^2_{jk}})^{-1}(\Lambda_{jk}-\Lambda_{0jk})^2 - \frac{1}{2}\kappa^2_{\Lambda_j}\tau^2_{\Lambda_{jk}})
\end{aligned}
$$

Multiplying and dividing $\left(\frac{\kappa^2_{\Lambda_j}}{(\Lambda_{jk}-\Lambda_{0jk})^2}\psi_{\epsilon j}\right)(\tau^2_{\Lambda_{jk}})^{-1}$, the exponent part of the last line becomes as follows.

$$
\begin{aligned}
&(-\frac{1}{2}\psi^{-1}_{\epsilon j}(\tau_{\Lambda^2_{jk}})^{-1}(\Lambda_{jk}-\Lambda_{0jk})^2 - \frac{1}{2}\kappa^2_{\Lambda_j}\tau^2_{\Lambda_{jk}}) \\
&= -\frac{\left(\psi^{-1}_{\epsilon j}(\tau_{\Lambda^2_{jk}})^{-1}(\Lambda_{jk}-\Lambda_{0jk})^2 + \kappa^2_{\Lambda_j}\tau^2_{\Lambda_{jk}}\right)\left(\frac{\kappa^2_{\Lambda_j}}{(\Lambda_{jk}-\Lambda_{0jk})^2}\psi_{\epsilon j}\right)(\tau^2_{\Lambda_{jk}})^{-1}}{2\left(\frac{\kappa^2_{\Lambda_j}}{(\Lambda_{jk}-\Lambda_{0jk})^2}\psi_{\epsilon j}\right)(\tau^2_{\Lambda_{jk}})^{-1}} \\
&= -\frac{\kappa^2_{\Lambda_j}\left((\tau^2_{\Lambda_{jk}})^{-2} + \frac{\kappa^2_{\Lambda_j}}{(\Lambda_{jk}-\Lambda_{0jk})^2}\psi_{\epsilon j}\right)}{2\left(\frac{\kappa^2_{\Lambda_j}}{(\Lambda_{jk}-\Lambda_{0jk})^2}\psi_{\epsilon j}\right)(\tau^2_{\Lambda_{jk}})^{-1}} \\
&= -\frac{\kappa^2_{\Lambda_j}\left((\tau^2_{\Lambda_{jk}})^{-2} - 2\tau^{-2}_{\Lambda_{jk}}\sqrt{\frac{\kappa^2_{\Lambda_j}}{(\Lambda_{jk}-\Lambda_{0jk})^2}\psi_{\epsilon j}} + \frac{\kappa^2_{\Lambda_j}}{(\Lambda_{jk}-\Lambda_{0jk})^2}\psi_{\epsilon j}\right) + 2\kappa^2_{\Lambda_j}\tau^{-2}_{\Lambda_{jk}}\sqrt{\frac{\kappa^2_{\Lambda_j}}{(\Lambda_{jk}-\Lambda_{0jk})^2}\psi_{\epsilon j}}}{2\left(\frac{\kappa^2_{\Lambda_j}}{(\Lambda_{jk}-\Lambda_{0jk})^2}\psi_{\epsilon j}\right)(\tau^2_{\Lambda_{jk}})^{-1}}
\end{aligned}
$$

Note that the second term of the numerator is actually a constant since $\tau^2_{\Lambda_{jk}}$ can be cancelled out. Therefore, this term can be ignored.

Now let $\mu_{\tau_{\Lambda_{jk}}} = \sqrt{\frac{\kappa^2_{\Lambda j}}{(\Lambda_{jk}-\Lambda_{0jk})^2}\psi_{\epsilon j}}$. Then $P(\tau^2_{\Lambda_{jk}}|\Lambda_{jk}, \psi_{\epsilon j}, \kappa^2_{\Lambda_j})$ can be expressed as follows.

$$P(\tau^2_{\Lambda_{jk}}|\Lambda_{jk}, \psi_{\epsilon j}, \kappa^2_{\Lambda_j}) \;\propto\; (\tau^2_{\Lambda_{jk}})^{-1/2} \exp\Big(- \frac{\kappa^2_{\Lambda_j}(\tau^{-2}_{\Lambda_{jk}} - \mu_{\tau_{\Lambda_{jk}}})^2}{2\mu^2_{\tau_{\Lambda_{jk}}}\tau^{-2}_{\Lambda_{jk}}}\Big)$$

Next, we derive $P(\tau^{-2}_{\Lambda_{jk}}|\Lambda_{jk}, \psi_{\epsilon j}, \kappa^2_{\Lambda_j})$ using 'change of variable'. At first, the determinant of Jacobian term can be computed as follows.

$$|J| = |\frac{d\tau^2_{\Lambda_{jk}}}{d\tau^{-2}_{\Lambda_{jk}}}| = |\frac{d(\tau^{-2}_{\Lambda_{jk}})^{-1}}{d(\tau^{-2}_{\Lambda_{jk}})}| = |-1 \times (\tau^{-2}_{\Lambda_{jk}})^{-2}| = \tau^4_{\Lambda_{jk}}$$

Combining this term with the above result,

$$\begin{aligned} P(\tau^{-2}_{\Lambda_{jk}}|\Lambda_{jk}, \psi_{\epsilon j}, \kappa^2_{\Lambda_j}) &\propto \tau^4_{\Lambda_{jk}}(\tau^2_{\Lambda_{jk}})^{-1/2}\exp\Big(- \frac{\kappa^2_{\Lambda_j}(\tau^{-2}_{\Lambda_{jk}} - \mu_{\tau_{\Lambda_{jk}}})^2}{2\mu^2_{\tau_{\Lambda_{jk}}}\tau^{-2}_{\Lambda_{jk}}}\Big) \\ &= \Big(\frac{1}{(\tau^{-2}_{\Lambda_{jk}})^3}\Big)^{1/2}\exp\Big(- \frac{\kappa^2_{\Lambda_j}(\tau^{-2}_{\Lambda_{jk}} - \mu_{\tau_{\Lambda_{jk}}})^2}{2\mu^2_{\tau_{\Lambda_{jk}}}\tau^{-2}_{\Lambda_{jk}}}\Big) \end{aligned}$$

The result indicates that the posterior distribution of $\tau^{-2}_{\Lambda_{jk}}$ is the 'Inverse-Gaussian' distribution introduced in Section 3.2.

$$\tau^{-2}_{\Lambda_{jk}}|\Lambda_{jk}, \psi_{\epsilon j}, \kappa^2_{\Lambda_j} \sim \text{IG}(\mu_{\tau_{\Lambda_{jk}}}, \kappa^2_{\Lambda_j})$$

ii) Posterior Distribution of $\kappa^2_{\Lambda_j}$

$$
\begin{aligned}
&P(\kappa^2_{\Lambda_j}|Y,\theta) \\
&\propto P(\tau_{\Lambda_j}|\kappa^2_{\Lambda_j})P(\kappa^2_{\Lambda_j}) \\
&\propto \prod_k^{q_{(j)}} \frac{\kappa^2_{\Lambda_j}}{2}\text{exp}(-\frac{1}{2}\kappa^2_{\Lambda_j}\tau^2_{\Lambda_{kl}}) \times (\kappa^2_{\Lambda_j})^{\alpha_{0\Lambda_j}-1}\text{exp}(-\beta_{0\Lambda_j}\kappa^2_{\Lambda_j}) \\
&\propto (\kappa^2_{\Lambda_j})^{\alpha_{0\Lambda_j}+q_{(j)}-1}\text{exp}(-\kappa^2_{\Lambda_j}(\beta_{0\Lambda_j}+\frac{1}{2}\sum_k^{q_{(j)}}\tau^2_{\Lambda_{jk}}))
\end{aligned}
$$

Hence,

$$
\begin{aligned}
\kappa^2_{\Lambda_j}|\tau_{\Lambda_j} &\sim \text{Gamma}(\alpha_{n\Lambda_j},\beta_{n\Lambda_j}) \\
\text{where } \alpha_{n\Lambda_j} &= \alpha_{0\Lambda_j}+q_{(j)}, \\
\beta_{n\Lambda_j} &= \beta_{0\Lambda_j}+\frac{1}{2}\sum_k^{q_{(j)}}\tau^2_{\Lambda_{jk}}
\end{aligned}
$$

It should be noted that the value of $\kappa_{\Lambda j}$ can be determined beforehand as in the Lasso, so that this prior and posterior are not needed.

Those above two parameters are concerned with regularization of the measurement model in SEM. Posterior distributions of their counterparts in the structural model, $\tau_{\Lambda_{\eta k}}$ and $\kappa^2_{\Lambda_{\eta k}}$ can be derived analogously, therefore omitted here.

Appendix D : R functions for Lasso SEM

In Chapter 4, we suggest the Double EM-algorithm for fitting Lasso SEM. The following is an R function that enables us to exploit the algorithm. Building this function, we referred 'SFAL1' function in Choi(2010).

```
library(lars)
lasso.sem <- function(data, lam.y = NULL, lam.x = NULL, constraint.y = NULL, constraint.x = NULL,
                      Beta = NULL, Gamma = NULL, lam.y0 = NULL, lam.x0 = NULL, Beta0 = NULL, Gamma0 = NULL,
                      phi0 = NULL, psi.y0 = NULL, psi.x0 = NULL, psi.eta0 = NULL, center = TRUE, scale = FALSE,
                      kappa.m = 1, kappa.s = 1, conv.gap = 0.001, num.iter = 1000,
                      normalize.m = TRUE, normalize.s = TRUE){

# lam.x, lam.y : Factor loading matrices specifying the measurement model.
#                'x' indicates measured variables for xi's(latent exogenous variables),
#                and 'y' indicates those for eta's(latent endogenous variables).
# constraint.y, constraint.x : Matrices whose rows indicates locations where the identification constraints are given.
#                              e.g., constraint.y <- rbind(c(1,1),c(4,2),c(7,3)); constraint.x <- rbind(c(1,1),c(4,2));
# Beta, Gamma : Coefficient matrices specifying the structural model.
# lam.y0, lam.x0, Beta0, Gamma0, phi0, psi.y0, psi.x0, psi.eta0 : Initial value matrices.
# center, scale : T/F. Indicates whether the user want to standardize the data or not.
# kappa.m : A tuning parameter for the measurement model.
# kappa.s : A tuning parameter for the structural model.
# conv.gap : The numeric value used as a criterion; if the absolute difference between [m-1]-th and [m]-th estimates
#            are less than conv.gap, we determines that the convergence occurs.
# num.iter : The maximum number of iteration.
# normalize.m : T/F. Indicates whether the regressors in the measurement model are normalized before the Lasso-fitting
#               algorithm is applied.
# normalize.s : T/F. Indicates whether the regressors in the structural model are normalized before the Lasso-fitting
#               algorithm is applied.


model <- list(lam.y, lam.x, Beta, Gamma)
lam0 <- list(lam.y0, lam.x0, Beta0, Gamma0)
# uniq0 <- list(psi.y0, psi.x0, psi.eta0)

p.y <- dim(lam.y)[1]; q1 <- dim(lam.y)[2];
p.x <- dim(lam.x)[1]; q2 <- dim(lam.x)[2];
p <- p.y + p.x; m <- q1 + q2;
# q1 : the number of endogenous latent variables
# q2 : the number of exogenous latent variables
# m : the number of latent variables (= q in the thesis)

#-------------------- Set model matrices

measurement <- matrix(0, p, m)
measurement[1:p.y, 1:q1] <- lam.y;
```

```
measurement[(p.y+1):p, (q1+1):m] <- lam.x

structural <- cbind(Beta, Gamma)


#-------------------- Checking errors --------------------

if(any(unlist(lapply(model, is.null)))){ cat("ERROR : The model was not specified. \n"); return(NULL);}
if(!any(unlist(lapply(model, is.null)))){
   if(is.null(constraint.y) | is.null(constraint.x)) {
      cat("ERROR : The function needs 'constraint',
          which indicates where the identification constraints will be given, when 'model' is specified. \n");
      return(NULL);}}


#-------------------- Set identification constraints --------------------

x.temp <- constraint.x
x.temp[,1] <- constraint.x[,1] + p.y
x.temp[,2] <- constraint.x[,2] + q1

constraint <- rbind(constraint.y, x.temp)


#-------------------- Set initial Values --------------------

if(any(unlist(lapply(lam0, is.null)))){
   lam.init.idx <- unlist(lapply(lam0, is.null));
   lam0[lam.init.idx] <- model[lam.init.idx];
}
lam.y0 <- lam0[[1]]; lam.x0 <- lam0[[2]];
Beta0 <- lam0[[3]]; Gamma0 <- lam0[[4]];

if(is.null(phi0)){phi0 <- matrix(0.25, q2, q2); diag(phi0) <- 0.5;}
if(is.null(psi.eta0)){psi.eta0 <- diag(0.3, q1)}
cfa.x <- lam.x0 %*% phi0 %*% t(lam.x0);
if(is.null(psi.x0)){psi.x0 <- diag(1 - diag(cfa.x));}
Iq <- diag(1, q1);
cov.eta <- solve(Iq-Beta0) %*% (Gamma0 %*% phi0 %*% t(Gamma0) + psi.eta0) %*% t(solve(Iq-Beta0));
cfa.y <- lam.y0 %*% cov.eta %*% t(lam.y0);
if(is.null(psi.y0)){psi.y0 <- diag(1 - diag(cfa.y));}

cov.eta.xi <- solve(Iq-Beta0) %*% (Gamma0 %*% phi0)


#-------------------- Load and check dataset --------------------

mv <- t(scale(data, center = center, scale = scale));  # p by n
n <- ncol(mv);
if(nrow(mv) != p){cat("ERROR : Wrong dataset! Check the number of variables! \n"); return(NULL);}
```

```
#-------------------- Some settings for iteration --------------------

count <- 0; # number of iterations

lam.y1 <- lam.y0; lam.x1 <- lam.x0;
lam1 <- matrix(0, p, m);
lam1[1:p.y, 1:q1] <- lam.y1;
lam1[(p.y+1):p, (q1+1):m] <- lam.x1

Beta1 <- Beta0; Gamma1 <- Gamma0;
lam.lv1 <- cbind(Beta1, Gamma1);

phi1 <- phi0;
psi.eta1 <- psi.eta0;
psi.y1 <- psi.y0; psi.x1 <- psi.x0;
psi1 <- diag(c(diag(psi.y1), diag(psi.x1)))

cov.eta1 <- cov.eta;
cov.eta.xi1 <- cov.eta.xi;
cov.xi.eta1 <- t(cov.eta.xi);
imp.lv1 <- rbind(cbind(cov.eta, cov.eta.xi1), cbind(cov.xi.eta1, phi0))
# implied covariance matrix of latent variables

imp.las <- lam1 %*% imp.lv1 %*% t(lam1) + psi1;
# implied covariance matrix

lam.y <- matrix(0, p.y, q1); lam.x  <- matrix(0, p.x, q2);
lam <- matrix(0, p, m);
lam.lv <- matrix(0, q1, m);
lam.lars <- matrix(0, p, m);
lam.lv.lars <- matrix(0, q1, m);


##############################################################################################
#-------------------- Iteration --------------------

while(max(abs(lam1-lam), abs(lam.lv1-lam.lv)) > conv.gap && count < num.iter )  {

   #-------------------- Reserve the estimates obtained in the previous loop --------------------

   lam   <- lam1;   lam.y <- lam.y1; lam.x <- lam.x1; phi     <- phi1;
   psi.y <- psi.y1; psi.x <- psi.x1; psi   <- psi1;   psi.eta <- psi.eta1;

   lam.lv <- lam.lv1; #cbind(Beta1, Gamma1)
   Beta   <- Beta1; Gamma  <- Gamma1;

   imp.lv <- imp.lv1; imp <- imp.las;
   imp.inv <- solve(imp);
```

```
############################################################
#-------------------- Measurement Model --------------------
############################################################

#------------------- Calculate delta and Delta --------------------

delta <- imp.inv %*% lam %*% imp.lv                                    # p by m
Delta <- imp.lv - imp.lv %*% t(lam) %*% imp.inv %*% lam %*% imp.lv     # m by m


#------------------- Calculate conditional expections of factor scores --------------------

exp.w <- t(delta) %*% mv                       # m by N, as factor scores.
exp.w.sq <- n * Delta + exp.w %*% t(exp.w)     # m by m, Sum of Wi's
Wmat <- exp.w.sq / n;                          # m by m, W matrix in the thesis

mv.mat <- mv %*% t(mv);
Cyy <- mv.mat/n;

# Temporary imp.lv
imp.lv.temp <- (1/2) * (Wmat + t(Wmat));  # It is based on Adachi(2013). This guarantees phi's symmetricity.
imp.lv.std <- diag(sqrt(1/diag(imp.lv.temp))) %*% imp.lv.temp %*% diag(sqrt(1/diag(imp.lv.temp)))
# Based on Rubin & Thayer(1982), and Adachi(2013);

#-------------------- Updating... --------------------
#----- Updating uniqueness(psi.y, psi.x) and lars estimates of Lambda

for(j in 1:p) {
   model.idx <- which(measurement[j, ] == 1)
   # This index indicates which factor scores are used as regressors for the measurement model.
   qj <- length(model.idx)
   # q(j) : The number of latent regressors for the j-th measurement variable.
   reg <- t(matrix(exp.w[model.idx, ], nrow = qj))
   # Regressor variables in the j-th measurement model equation.
   resp <- mv[j,]
   # Response variable in the j-th measurement model equation.

   #----- Updating uniqueness
   # Choi(2010)'s uniqueness estimates for CFA
   psi1[j,j] <- (1/n) * (mv.mat[j, j] - 2*(mv.mat[j,]  %*% delta[, model.idx] %*% lam[j, model.idx])
                         + t(lam[j, model.idx])%*%(exp.w.sq[model.idx, model.idx])%*%lam[j, model.idx]);
   if(psi1[j,j] < 0.005) psi1[j,j]<-0.005
   # Set 0.005 as a lower bound of uniqueness
   # This will prevent 'Non positive uniqueness' problem.

   #----- Updating lam.lars
   # psi correction + Cholesky decompositiom

   delta.j  <- delta[, model.idx]
   Delta.j  <- Delta[model.idx, model.idx]
```

```
   chol.mat <- matrix(Delta.j + t(delta.j) %*% (Cyy) %*% delta.j, nrow = qj, ncol = qj)
   # W(j) matrix in Cholesky decomposition

   # Standardiztion of regressor for Lasso estimation
   if(normalize.m) {
      reg <- scale(reg);
      chol.mat <- diag(1/sqrt(diag(chol.mat)), nrow = qj) %*% chol.mat %*% diag(1/sqrt(diag(chol.mat)), nrow = qj)
      resp <- t(scale(resp, center = TRUE, scale = FALSE));
   }
   chol.mat <- n * chol.mat
   X.tilde <- chol(chol.mat / psi1[j,j])                      # Regressor input of LARS function

   # In case we use SVD instead of Cholesky decomposition
   # X.svd <- svd(chol.mat / psi[j,j])
   # X.tilde <-  t(X.svd$u %*% diag(x=sqrt(X.svd$d), nrow = length(X.svd$d)))

   y.tilde <- t( resp %*% reg %*% solve(X.tilde) / psi1[j,j])  # Response input of LARS function

   # Fitting LARS
   b <- lars(X.tilde, y.tilde, intercept = FALSE, normalize = FALSE)
   lam.lars[j, model.idx] <- coef(b, s = kappa.m, mode = "lambda")

}
psi.y1 <- psi1[1:p.y, 1:p.y]
psi.x1 <- psi1[(p.y+1):p, (p.y+1):p]

#----- Compute new lambda estimates using lam.lars

cons <- matrix(constraint, nrow = m)
lam.scaling.mat <- diag(1/lam.lars[cons])
imp.lv.scaling.mat <- diag(lam.lars[cons])

# unstandardized lambda
lam1 <- lam.lars %*% lam.scaling.mat
lam.y1 <- lam1[1:p.y, 1:q1]
lam.x1 <- lam1[(p.y+1):p, (q1+1):m]


############################################################
#-------------------- Structural Model --------------------
############################################################

#------------------- Calculate conditional expections of factor scores --------------------

lv.eta <- t(exp.w[1:q1, ])
lv.xi  <- t(exp.w[(q1+1):m, ])

#-------------------- Updating... --------------------
#----- Updating residual variance of latent variables(psi.eta), and lars estimates of Beta and Gamma

for(k in 1:q1) {
```

```
model.idx <- which(structural[k, ] == 1)
# This index indicates which factor scores are used as regressors for the structural model.
qk <- length(model.idx)
# q(k) : The number of latent regressors for the k-th equation in the structural model.
lv.reg <- t(matrix(exp.w[model.idx, ], nrow = qk))
# Regressor variables in the k-th equation in the structural model.
lv.resp <- exp.w[k, ]
# Response variable in the k-th equation in the structural model.

#----- Updating residual variance, psi.eta
term1 <- Delta[k, k] + t(delta[,k]) %*% Cyy %*% delta[,k]
term2 <- 2 * lam.lv1[k, model.idx] %*% ( Delta[model.idx, k] + t(delta[,model.idx]) %*% Cyy %*% (delta[,k]) )
term3 <- lam.lv1[k, model.idx] %*%
   (Delta[model.idx, model.idx] + t(delta[,model.idx]) %*% Cyy %*% delta[,model.idx]) %*% t(t(lam.lv1[k, model.idx]))
psi.eta1[k,k] <- term1 - term2 + term3

if(psi.eta1[k,k] < 0.005) psi.eta1[k,k]<-0.005
# Set 0.005 as a lower bound of residual variance
# This will prevent 'non positive residual variance' case.


#----- Updating lars estimates
# psi correction + Cholesky decomposition
chol.lv.mat0 <- matrix(Wmat[model.idx, model.idx], nrow = qk, ncol = qk)
# WAk matrix =  W[1:q(k), 1:q(k)]
Vkmat0 <- Wmat[c(k, model.idx), c(k, model.idx)]
# W[c(k, 1:q(k)), c(k, 1:q(k))] ; this matrix is used for obtaining V_{eta,k vector}

# Standardiztion of regressor for Lasso estimation
if(normalize.s) {
   lv.reg <- scale(lv.reg);

   Wqk.scale  <- diag(1/sqrt(diag(chol.lv.mat0)), nrow = qk)
   Vk.scale <- diag(c(1, diag(Wqk.scale)));

   chol.lv.mat0 <- Wqk.scale %*% chol.lv.mat0 %*% Wqk.scale
   Vkmat0 <- Vk.scale %*% Vkmat0 %*% Vk.scale
}

chol.lv.mat <- n * chol.lv.mat0
Omega.tilde <- chol(chol.lv.mat / psi.eta1[k,k])                    # Regressor input of LARS function

# In case we use SVD instead of Cholesky decomposition
# Omega.svd <- svd(chol.lv.mat / psi.eta1[k,k])
# Omega.tilde <-  t(Omega.svd$u %*% diag(x=sqrt(Omega.svd$d), nrow = length(Omega.svd$d)))

Vkmat <- n * Vkmat0
V.eta.k <- Vkmat[1, 2:(1+qk)]
eta.tilde.lv <- t( V.eta.k %*% solve(Omega.tilde) / psi.eta1[k,k])  # Response input of LARS function

# Fitting LARS
```

```
    b <- lars(Omega.tilde, eta.tilde.lv, intercept = FALSE, normalize = FALSE)
    lam.lv.lars[k, model.idx] <- coef(b, s = kappa.s, mode = "lambda")


    ##############################################################################################################
    # We can use MM-algorithm in estimating Lasso SEM, instead of LARS.
    # MM can be applied in various ways. The followings are some examples.
    #
    #----- MM-algorithm with conditional expectation of log likelihood (using linearity only)
    #if(normalize.s) {lv.resp <- scale(lv.resp, center = TRUE, scale = FALSE); lv.reg <- scale(lv.reg)}
    #pre.coef <- lam.lv[k, model.idx]
    #Vvec <- n * (Delta[model.idx, k] + t(delta[, model.idx]) %*% Cyy %*% delta[,k])
    #Wmat <- n * (Delta[model.idx, model.idx] + t(delta[, model.idx]) %*% Cyy %*% delta[,model.idx])
    #lam.lv1[k, model.idx] <- (Vvec/psi.eta1[k,k]) / (rowSums(Wmat)/psi.eta1[k,k] + (1/2) * (kappa.s/abs(pre.coef)))
    #
    #----- MM-algorithm with conditional expectation of log likelihood (using convexity and linearity)
    #if(normalize.s) {lv.resp <- scale(lv.resp, center = TRUE, scale = FALSE); lv.reg <- scale(lv.reg)}
    #pre.coef <- lam.lv[k, model.idx]
    #alpha <- abs(lv.reg) / rowSums(abs(lv.reg))
    #nom1 <- t(lv.reg) %*% (lv.resp - lv.reg %*% pre.coef);
    #nom2 <- diag(t(lv.reg / alpha) %*% lv.reg) * pre.coef
    #denom1 <- diag(t(lv.reg / alpha) %*% lv.reg) / psi.eta1[k,k]
    #denom2 <- (1/2) * kappa.s / abs(pre.coef)
    #lam.lv1[k, model.idx] <- 1/psi.eta1[k,k] * (nom1 + nom2) / (denom1 + denom2)
    ##############################################################################################################
}

# Unstandardized lambda of latent variable
lam.lv1 <- lam.lv.lars
Beta1  <- lam.lv1[, 1:q1];
Gamma1 <- lam.lv1[, (q1+1):m];

#----- Compute covariance / correlation of latent variables

# LARS scaling for the var-cov matrix, in order to obtain xi's covariance matrix Phi
#   (using measurement model's lars result)
imp.lv.lars <- imp.lv.scaling.mat %*% imp.lv.std %*% imp.lv.scaling.mat
# Implied latent covariance matrix scaled by lars estimates
phi1 <- imp.lv.lars[(q1+1):m, (q1+1):m]
# covariance of xi's

# Compute the var-cov matrix of eta's and covariance between eta's and xi's
cov.eta1 <- solve(Iq-Beta1) %*% (Gamma1 %*% phi1 %*% t(Gamma1) + psi.eta1) %*% t(solve(Iq-Beta1));
cov.eta.xi1 <- solve(Iq-Beta1) %*% (Gamma1 %*% phi1)
cov.xi.eta1 <- t(cov.eta.xi1);

# Implied covariance matrix of latent variables
imp.lv1 <- rbind(cbind(cov.eta1, cov.eta.xi1), cbind(cov.xi.eta1, phi1))

# implied covariance matrix
imp.las <- lam1 %*% imp.lv1 %*% t(lam1) + psi1
```

```
   #----- end of updating estimates

   count <- count + 1;

      } #----- end of iteration

      # standardized lambda
      std.lam <- diag(1/sqrt(diag(imp.las))) %*% lam1 %*% diag(sqrt(diag(imp.lv1)))
      std.lam.y <- std.lam[1:p.y, 1:q1]
      std.lam.x <- std.lam[(p.y+1):p, (q1+1):m]

      # standardized Beta and Gamma
      std.lam.lv <- diag(1/sqrt(diag(cov.eta1))) %*% lam.lv1 %*% diag(sqrt(diag(imp.lv1)))
      std.Beta <- std.lam.lv[, 1:q1];
      std.Gamma <- std.lam.lv[, (q1+1):m];

      #-------------------- Output --------------------
      output <- list(lam.lars = lam.lars, lam.lv.lars = lam.lv.lars, lambda.y = lam.y1, lambda.x = lam.x1,
                     std.lambda.y = std.lam.y, std.lambda.x = std.lam.x, Beta = Beta1, Gamma = Gamma1,
                     std.Beta = std.Beta, std.Gamma = std.Gamma, psi.y = psi.y1, psi.x = psi.x1, psi.eta = psi.eta1,
                     phi = phi1, imp = imp.las, imp.lv = imp.lv1, iteration = count);
      output;
}
```

Appendix E : Generating Population in SEM

In Chapter 1 and 5, we discussed an issue regarding the conceptualization of population covariance matrix in Structural Equation Modeling. In order to simulate the true data-generating process, we concluded that the conceptualization c) should be exploited. This implies that the population is generated from the following covariance matrix, which consists of a systematic function component and a perturbation component.

$$\Sigma_0^* = \Sigma_0(\theta_0) + E$$

In this section, we shall introduce the 'Cudeck-Browne Procedure', which is proposed by Cudeck and Browne(1992) for the purpose of deriving the matrix Σ_0^* appropriately to meet some requirements. As we described in the previous chapters, this conceptualization considers model error by adding the matrix E. Note that this inclusion is carried out with special care to satisfy several conditions.

i) $\underset{\theta}{\text{argmin}}(F(\Sigma_0^*, \Sigma_0(\theta))) = \theta_0$.
ii) For some predetermined value of δ, $F(\Sigma_0^*, \Sigma_0(\theta_0)) = \delta$.

where $F(\cdot, \cdot)$ indicates the discrepancy function described in Section 1.2. In general, this function can be expressed as follows.

$$F(\Sigma_0^*, \Sigma(\theta)) = \frac{1}{2}\text{tr}([W^{-1}(\Sigma_0^* - \Sigma(\theta))]^2)$$

When the maximum likelihood principle is exploited to estimate SEM, F is described in slightly different form.

$$F_{ML}(\Sigma_0^*, \Sigma(\theta)) = \log|\Sigma(\theta)| - \log|\Sigma_0^*| + \text{tr}(\Sigma_0^*\Sigma(\theta)^{-1}) - p$$

Note that this function is minimized at the same point with $F(\Sigma_0^*, \Sigma(\theta))$ defined with $W = \Sigma(\theta_{ML})$, but the value of minimized function is different from each other(Cudeck & Browne, 1992).

The condition i) and ii) represent that, the perturbation matrix E must not affect the population value of θ_0 and degree of model error can be adjusted by assigning appropriate value of δ. With this requirements satisfied, researchers are able to study SEM by Monte Carlo experiments considering proper model error without being interfered in retrieving the parameter value they pre-specified (But note that correct retrieval occurs only when the model function is correctly specified as $\Sigma_0(\cdot)$).

The Cudeck-Browne Procedure can be introduced briefly as follows. Definitions of some operators and terms are drawn from their original paper and other references(Browne, 1973; Nel, 1980; Magnus & Neudecker, 1999).

1) Define vec(·) and vecs(·) operators.

2) Define K_p, the transition matrix satisfying $\text{vecs}(A) = K_p^T \text{vec}(A)$.

3) Compute $K_p^- = (K_p^T K_p)^{-1} K_p^T$, the Moore-Penrose inverse of K_p. This is an inverse transition matrix satisfying $\text{vec}(A) = (K_p^-)^T \text{vecs}(A)$.

4) Define $D_k = (K_p^-)(K_p^-)^T$.

5) Compute the partial derivative of model function $\Sigma_0(\cdot)$ with respect to each of parameters.

$$\dot{\Sigma}_i = \partial \Sigma_0(\theta) / \partial \theta_i, \quad i = 1, \cdots, q.$$

q : number of parameters

This can be replaced with the forward difference approximation.

$$\dot{\Sigma}_i \doteqdot \frac{\Sigma_0(\theta + hu_i) - \Sigma_0(\theta)}{h}$$

where u_i is an unit vector having only one non-zero element, which is 1, in i-th position like standard basis. And h is an appropriately small number, such as $h = 10^{-8}$.

6) Compute the partial derivative of the discrepancy function F with respect to each of parameters.

$$\begin{aligned}
\frac{\partial F(\Sigma_0^*, \Sigma_0(\theta))}{\partial \theta_i} &= -\text{tr}((\Sigma_0^* - \Sigma_0(\theta))W^{-1}\dot{\Sigma}_i W^{-1}) \\
&= -\text{vecs}(W^{-1}\dot{\Sigma}_i W^{-1})^T (K_p^-)(K_p^-)^T \text{vecs}(\Sigma_0^* - \Sigma_0(\theta)) \\
&= -\text{vecs}(W^{-1}\dot{\Sigma}_i W^{-1})^T D_k \text{vecs}(\Sigma_0^* - \Sigma_0(\theta)) \\
&= b_i^T \text{vecs}(\Sigma_0^* - \Sigma_0(\theta))
\end{aligned}$$

where $b_i = -D_k \text{vecs}(W^{-1}\dot{\Sigma}_i W^{-1})$ and W is a weight matrix in function F.

7) Compute the vector b_i's and define the matrix B = ($\text{b}_1, b_2, \cdots, b_q$).

By the step 1-7, we obtain the gradient of F.

$$\frac{\partial F(\Sigma_0^*, \Sigma_0(\theta))}{\partial \theta} = B^T \text{vecs}(\Sigma_0^* - \Sigma_0(\theta)) = B^T \tilde{e}$$

where $\tilde{e} = \text{vecs}(\Sigma_0^* - \Sigma_0(\theta))$.

8) For the condition i), the following first order condition should be satisfied.

$$B^T \tilde{e} = 0 \mid_{\theta=\theta_0}$$

In order to obtain an appropriate vector $\tilde{e}$, consider $y = Bv + \epsilon$, where

y is an arbitrary vector with proper length. By the OLS estimation, we obtain $\hat{v} = (B^TB)^{-1}B^Ty$ and this yields $\tilde{e} \overset{\text{def}}{\equiv} \hat{\epsilon} = y - B\hat{v} = (I - H)y$, where $H = B(B^TB)^{-1}B^T$. This satisfies $B^T\tilde{e} = B^T(I - H)y = 0$.

9) In general, choice of y is not important. However, when the OLS method is exploited, this can be the issue since some y's may produce non-positive definite Σ_0^*. With respect to this point, Cudeck and Browne(1992) suggested the following derivation of y.

i) Define $A_1 = U^TU$, where U_i is an $m \times p(m > p)$ matrix consisting of random samples from the standard uniform distribution. This yields at least p.s.d. matrix A_1.

ii) Rescale A_1 to obtain A_2 ; $A_2 = diag(A_1)^{-1/2}A_1diag(A_1)^{-1/2}$. This yields at least p.s.d. matrix with unit-diagonal.

iii) In the model with the measurement part such as FA and standard SEM, Rescale $A_3 = {\Psi_\epsilon}^{1/2}A_2{\Psi_\epsilon}^{1/2}$. This yields $Diag(A_3) = \Psi_\epsilon$.

iv) Finally, obtain $y = \text{vecs}(A_3)$. This y can produce a positive definite Σ_0^* when it is used in Cudeck-Browne Procedure.

10) For the condition ii), define $E = \kappa\tilde{E}$. Here, κ is a coefficient making F function produce δ when it is minimized. And finding appropriate value of κ, we can compute $\Sigma_0^* = \Sigma_0(\theta_0) + E$. Note that since $B^T\tilde{e} = 0$, it is also satisfied that $B^Te = B^T(\kappa\tilde{e}) = 0$. Moreover, Browne and Cudeck mentioned that with not-too-much large value of κ, it is guaranteed to obtain the global minimizer of F.

The value of κ can be computed as in the following manner. With prespecified δ,

i) In general case, let $G = W^{-1}\tilde{E}$. Then,

$$\begin{aligned}\delta &= F(\Sigma_0^*, \Sigma_0(\theta)) \\ &= \frac{1}{2}\text{tr}([W^{-1}(\Sigma_0^* - \Sigma_0(\theta))]^2) = \frac{1}{2}\text{tr}((\kappa G)^2) \\ &\Rightarrow \kappa = [2\delta/\text{tr}(G^2)]^{1/2}\end{aligned}$$

ii) In case we exploit the ML method for estimation,

$$G = W^{-1}\tilde{E} = \Sigma_0(\theta_0)^{-1}\tilde{E} \quad \Rightarrow \quad \Sigma_0(\theta_0)^{-1}E = \Sigma_0(\theta_0)(\kappa\tilde{E}) = \kappa G$$

$$\begin{aligned}\min_\theta[F_{ML}(\Sigma_0^*, \Sigma_0(\theta))] &= F_{ML}(\Sigma_0^*, \Sigma_0(\theta_0)) \\ &= \log|\Sigma_0(\theta_0)| - \log|\Sigma_0(\theta_0) + E| + \text{tr}((\Sigma_0(\theta_0) + E)\Sigma_0(\theta_0)^{-1}) - p \\ &= -\log\frac{|\Sigma_0(\theta_0) + E|}{|\Sigma_0(\theta_0)|} + \text{tr}(I_p + \Sigma_0(\theta_0)^{-1}E) + p \\ &= -\log|I_p + \Sigma_0(\theta_0)^{-1}E| + \text{tr}(\Sigma_0(\theta_0)^{-1}E) \\ &= -\log|I_p + \kappa G| + \kappa\ \text{tr}(G)\end{aligned}$$

This minima should be equal to, or at least close to δ so as to satisfy the condition ii). But this problem has no closed form solution. Thus some numerical methods should be exploited. Cudeck and Browne(1992) suggest to use Newton-Rhapson method to solve the equation. At first, define the following function and derive its derivative.

$$\begin{aligned}t(\kappa) &= F_{ML}(\Sigma_0^*, \Sigma_0(\theta_0)) - \delta \\ &= \kappa\ \text{tr}(G) - \log|I_p + \kappa G| - \delta \\ t'(\kappa) &= \text{tr}(G) - \text{tr}[(I_p + \kappa G)^{-1}G]\end{aligned}$$

Our goal is to minimize $t(\kappa)$ so that the value of the function becomes as close to zero as possible. This makes the minima of F_{ML} approximately equal

to δ. Using the above functions, Newton-Rhapson algorithm yields the iterative solution by repeating

$$\kappa_{n+1} = \kappa_n - \frac{t(\kappa_n)}{t'(\kappa_n)}$$

Initial value κ_0 can be obtained by the result of general case, $\kappa = [2\delta/\text{tr}(G^2)]^{1/2}$, as Cudeck and Browne suggested.

With the above procedure, we can obtain Σ_0^*. And when we try to minimize the discrepancy function, the result will retrieve the solution θ_0 and F-value is computed to be δ.

초록

심리학 연구의 일반화가능성/재현가능성 문제가 화두에 오른 현재, 이를 다루고 있는 연구와 논의들은 대부분 인센티브 구조, 실험의 엄밀성 부족, 편파적인 결과 보고 등 '실험 혹은 절차'의 맥락에서 문제에 접근하고 있다. 그러나 기존에 널리 사용되는 통계분석 기법들에도 연구의 일반화가능성/재현가능성을 제한하는 특징들이 내재되어 있는 경우가 존재한다. 본고에서는 이러한 점에 착안하여 '분석 혹은 통계 기법'의 맥락에서 문제에 접근하고자 하였다. 이를 위해 심리학에서 널리 사용되는 통계 모형인 구조방정식모형에 대한 개선책이 연구되었다. 즉 기존의 최대가능도추정법보다 분석 결과의 변산성을 줄일 수 있는 기법으로 회귀모형의 Regularization, 그 중에서도 Lasso를 구조방정식모형에 적용하고자 하였다.

우선 일반화가능성/재현가능성을 표상할 수 있는 지수로서 Overall Discrepancy와 Mean Squared Error 등에 대해 논의하였다. 그리고 구조방정식모형에 대해 Lasso를 적용하고자 한 선행 연구로서, 최근 생물학 분야에서 제시된 Bayesian Lasso SEM에 대해 연구하고 근본적인 한계점을 지적하였다. 나아가 본고에서는 구조방정식모형에 Lasso를 직접 적용하는 방법으로서 'Double EM-algorithm'을 제안하고 시뮬레이션 연구를 통해 그 성능에 대해 조사하였다.

시뮬레이션 연구는 요인분석모형과 구조방정식모형에 대해, 의도적으로 잘못 투입된 여러 경로들을 상정하여 실시하였다. Lasso SEM를 통해 이러한 경로들을 제거하고 모집단 생성 과정에 가까운 모형을 회복할 수 있는지, 그리고 이를 통해 일반화가능성 지수들을 개선할 수 있는지를 살펴보는 것이 가장 큰 목적이었다. 동시에 모형 오차, 표본의 크기, 공분산 행렬의 값의 크기 등 다양한 조건이 고려되어 Lasso SEM가 어떤 조건에서 좋은 결과를 도출해낼 수 있는지를 살펴보고자 하였다.

다양한 조건에서 분석에 임한 결과, Lasso SEM이 여러 조건에서 기존의 방법보다 여러 일반화가능성 지수를 개선할 수 있음이 밝혀졌다. 동시에 모형에 의도적

으로 잘못 투입된 경로들을 탐지하여 제거하는 효과도 검증되었다. 그러나 이러한 성능들은 조건에 따라 달리 발휘되어, Lasso SEM의 실제 사용은 다루고자 하는 자료의 특성에 따라 다른 결과를 도출할 수 있음을 시사하였다. 특히 모집단 생성 과정에 포함되는 모형 오차는 구조방정식모형에서 Lasso의 성능을 저해할 수 있는 가장 강력한 요인으로 규명되었다. 저자는 이러한 문제가 Overall Discrepancy에 입각한 최적화 방식에 있음을 지적하고, 그 대안으로 최적화의 기준이 되는 지수나 목적함수를 교정할 것을 제안하였다. 이는 Lasso SEM의 모형 오차에 대한 취약성을 개선할 뿐만 아니라 계수 추정의 정밀성을 높일 수 있을 것이다.

동시에 실시된 상관 연구 결과에서는 기존의 추정과정에서 사용되는 Sample Discrepancy, 그리고 그에 입각하여 정의된 여러 모형적합도 지수들이 Overall Discrepancy, Mean Squared Error 등의 일반화가능성 지수와 굉장히 낮은 상관을 보인다는 점이 발견되었다. 이는 전통적인 방법으로 추정한 구조방정식모형의 결과물이 미래의 자료, 그리고 해당 모형이 설명하고자 하는 실제 현상에 일반화되기 어려울 수 있다는 점을 시사하였다.

주요어: 재현 가능성, 일반화 가능성, 구조방정식모형, 요인분석모형, 정규화, 벌점화, Lasso, Overall Discrepancy, Mean Squared Error
학번: 2014-20229